

# The E2 Proteins

Alison McBride<sup>a</sup> and Gerald Myers<sup>b</sup>

<sup>a</sup>Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

<sup>b</sup>Theoretical Biology and Biophysics, T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

## Introduction

The papillomavirus E2 proteins regulate viral transcription and replication and therefore play a central role in the viral life-cycle. In BPV, the full-length E2 gene product is a transcriptional transactivator that activates transcription from several viral promoters by binding to E2 binding sites located within enhancer elements in the LCR (reviewed in [57,61]). BPV also encodes two shorter forms of the E2 protein which antagonize the function of the full-length transactivator. The E2 repressors inhibit transcription by binding to and blocking the E2 binding sites and/or by forming heterodimers with the E2 transactivator. The E2 proteins of the human papillomaviruses can also function as transcriptional transactivators but most, though not all, studies find that they repress the activity of the E6/E7 gene promoter [6,9,67,78].

Viral DNA replication requires the full-length E2 transactivator, the viral E1 protein and the replication origin. The replication origin contains an E1 binding site flanked on either side by E2 binding sites. The E1 protein has several replication-associated activities such as origin-specific binding and helicase activities and forms a complex with the E2 transactivator. The E2 protein probably plays an auxiliary role in replication by enhancing and regulating the functions of the E1 protein.

Plasmids containing the minimal replication origin can replicate transiently in cells expressing the E1 and E2 proteins but with time the replicated DNA is lost. Long term, stable maintenance of such plasmids requires additional E2 binding sites and expression of the E1 and E2 proteins [63]. Therefore, it appears that the E2 protein may also play a role in plasmid copy number control and viral genome segregation.

The E2 proteins may also play a role in packaging the viral genomes in virion particles. Viral DNA appears to be packaged much more efficiently in the presence of the E2 protein, both in insect and mammalian cell lines [70,90].

Thus, the E2 proteins are multifunctional and important for several steps of the viral life cycle. Most of the knowledge about the structure and function of the E2 proteins has been obtained with the BPV1 E2 proteins and this review will concentrate on these proteins, unless otherwise stated. However, in general, the structure and functions of the E2 proteins seem to be comparable in all E2 proteins that have been examined to date. This review will summarize studies that have mapped functions to the various regions of the E2 proteins

## A. E2 gene products

The E2 proteins have been best characterized for BPV1. Three BPV1 E2 proteins have been identified and mapped to the E2 ORF [36,39]. The largest 48kD protein, expressed from the entire ORF, is a transcriptional transactivator and is required for viral DNA replication [73,81]. This protein has been designated E2-TA. Two smaller proteins, encoded by the 3' half of the ORF, function as transcriptional repressors [17,41]: E2-TR is a 30kD protein expressed from the P3080 promoter and initiated at an internal initiation codon at residue 162; E8/E2 is a 28kD protein encoded by a message with a 1234 $\wedge$ 3225 splice which encodes 11 amino acids from the E8 ORF joined to the C-terminal 205 amino acids of E2. HPV cDNAs that are capable of encoding C-terminal regions of HPV E2 proteins have been identified [3,15,25] but, as yet, there is no direct genetic or biochemical evidence that the human papillomaviruses encode truncated E2 repressor proteins.

## E2 Proteins

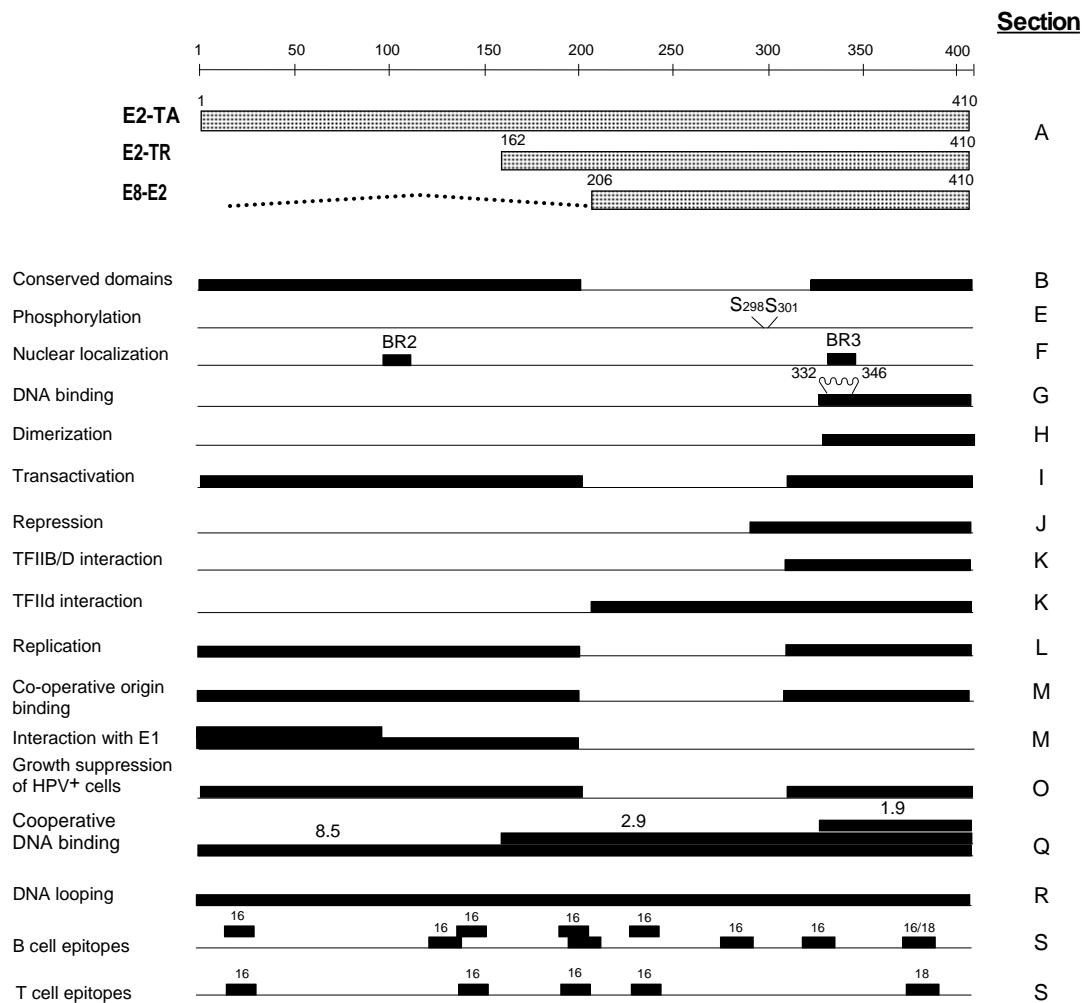


Figure 1. The structure of the three BPV1 E2 proteins are shown at the top of the figure. Below, the functions that have been mapped to different regions of the proteins are indicated. Refer to the section indicated to the right for more details. An alignment of E2 amino acid sequences is presented in Appendix A.

### B. Conserved domains

Analysis of the predicted amino acid sequence of all papillomavirus E2 proteins shows that there are two conserved domains (see figure 1 and appendix A). An N-terminal domain of about 200 amino acids and a C-terminal domain of about 90 amino acids are separated by a non-conserved region of variable length that has been designated the “hinge” region. Notably, the hinge region overlaps the E4 open reading frame which is quite divergent among the papillomaviruses. The conserved E2 domains are approximately 35% similar among the papillomaviruses (Appendix C).

### C. Protein structure

The E2-TA polypeptide consists of a C-terminal DNA binding domain linked to an N-terminal transactivation domain by a non-conserved hinge region. The E2 protein forms dimers that are mediated through the DNA binding domain and, as described below, the structure of the C-terminal domain has been solved.

The transactivation domain is approximately 200 amino acids and, unlike many other transactivation domains, appears to have a very constrained structure that is easily disrupted by deletion or certain non-conservative point mutations. The amino acid sequence of almost all of the papillomavirus E2 proteins is predicted to form two  $\alpha$ -helices in the N-terminal region of the transactivation domain ([29] and Appendix B). However, as yet, there is no experimental evidence that such secondary structures exist in the transactivation domain.

The hinge region of the E2 proteins varies both in length and in amino acid composition among the E2 proteins. It has been postulated that this region forms a flexible link between the two domains and a study of the HPV16 E2 protein confirmed that the hinge is an unstructured region [28]. Antibodies were generated against overlapping peptides covering the entire E2 protein and it was found that only antibodies against the hinge region can recognize the native, undenatured E2 protein.



Figure 2. X-ray crystal structure of BPV-1 DNA binding domain (326–410) bound to DNA [33].

#### D. Protein turnover and Cell cycle Expression

The relative ratios of the three BPV1 E2 proteins have been measured in virally-transformed C127 cells as 1:10:3 for E2-TA/E2-TR/E8-E2 [36]. Within these cells E2-TA has a half-life of approximately 40 minutes and E2-TR and E8-E2 have half-lives of 10 and 15 minutes, respectively [36]. The ratio of the three BPV1 E2 proteins changes throughout the cell cycle with the ratio of E2 transactivator to repressors being highest at S phase and lowest at G1 [89].

#### E. Phosphorylation

BPV1, CRPV, HPV11 and HPV16 E2 proteins have been shown to be phosphorylated [4,10,53, 58,69] but the phosphorylation sites have only been identified in BPV1 E2 [53]. BPV1 E2 contains both phosphoserine and phosphothreonine but only the two major serine phosphorylation sites at positions 298 and 301 have been mapped [53]. Mutation of E2 serine 301 to alanine results in a virus that replicates to a much greater copy number than wildtype BPV1 [55]. However, this region of the BPV1 E2 protein is not conserved among the other papillomavirus E2 proteins.

## E2 Proteins

### F. E2 localization

All three BPV1 E2 proteins are located in the nucleus but a greater percentage of the full-length E2-TA protein is associated with insoluble chromatin and nuclear matrix components [36]. Two putative nuclear localization signals (NLS) have been identified in the BPV1 E2 proteins. A C-terminal peptide (residues 339–352, KCYRFRVKKNHRHR) which contains the DNA recognition helix of the DNA binding domain functions as a NLS both in the DNA binding domain and in heterologous proteins [71]. Deletion or mutation of a second signal in the transactivation domain (residues 107 to 115, KRCFKKGAR) results in a cytoplasmic E2 protein even though the C-terminal NLS is present. Therefore, it has been postulated that C-terminal NLSs may be masked in the E2-TA protein [71]. A recent study has shown that point mutations in the BR2 region (K111A, K112A) cause the protein to oligomerize and be retained in the cytoplasm [1].

High levels of E2 expression are found primarily in the stratum spinosum of infected wart tissue which coincides with the region in which viral genome amplification occurs [13]. This may indicate that high levels of E2 are important for the switch to vegetative viral DNA replication.

### G. DNA binding

The C-terminal domain of E2 (residues 326–410) binds specifically to DNA as a dimer (reviewed in [57]). The consensus E2 recognition sequence is ACCN<sub>6</sub> GGT but the internal nucleotides and flanking sequences can greatly influence the affinity of E2 binding over several orders of magnitude [47]. The X-ray crystal structure of the C-terminal 85 amino acids of E2 bound to DNA was the first example of an anti-parallel  $\beta$ -barrel DNA binding structure [33]. As shown in figure 2, a dimer of the E2 DNA binding domain forms an eight-stranded anti-parallel  $\beta$ -barrel made up of four strands from each subunit. A pair of  $\alpha$ -helices symmetrically positioned on the outside of the barrel contain the amino acids residues that are required for specific DNA interaction. Figure 3 shows the amino acid sequence of this recognition helix aligned with the homologous region from other papilloma virus E2 proteins. Also indicated on the figure are specific mutations that have been generated in this region of the E2 protein and their effect on DNA binding. The three dimensional structure of the HPV31 DNA binding domain in solution has also been determined by nuclear magnetic resonance [50]. The overall protein fold is very similar to the crystal structure of the BPV-1 domain but the DNA recognition helix appears to be flexible, as has been observed in a number of other DNA binding proteins. The DNA binding domain of the Epstein Barr virus EBNA1 protein has a very similar structure to the E2 DNA binding domain despite no sequence similarity [8].

The DNA recognition helix contains a highly conserved cysteine residue at position 340 that is very sensitive to oxidation [56]. This residue makes direct contacts with DNA yet E2 proteins with certain substitutions of this residue (glycine, serine, alanine) are still able to bind DNA. However, these proteins are not able to activate transcription efficiently in mammalian cells [31,56]. Similar reactive cysteines are found in the basic DNA-recognition regions of other proteins such as fos, jun and NF B.

### H. Dimerization

The DNA binding domain of the E2 proteins forms a stable dimer even in the absence of DNA [18,54]. The DNA binding and dimerization properties of this domain cannot be separated by deletion analysis; all deletions that have been tested eliminate both properties of the C-terminal domain. This is because dimerization involves an extensive subunit interface consisting of inter-backbone hydrogen bonds between the  $\beta$ -strands, interaction between side-chains of the  $\beta$ -strands in each subunit and an extensive hydrophobic core. This hydrophobic core contains a highly conserved tryptophan residue at position 360 which has been designated the tryptophan bridge ([18,64] see figure 2). The indole rings of W360 from each subunit are in Van der Waals contact which allows them to be crosslinked by UV irradiation [64]. Mutated E2 proteins containing hydrophobic residues at this position are functional but substitution of W360 by polar residues disrupts dimerization. Non-conservative mutations in other parts of the C-terminal domain can also eliminate both DNA binding and dimerization by disrupting protein structure [14].

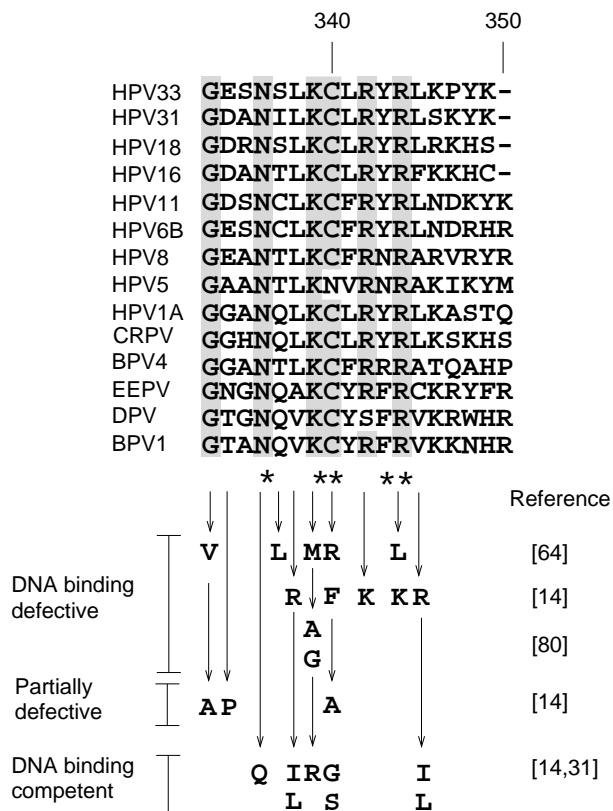


Figure 3. Alignment of the amino acid sequence of the DNA recognition helix of several papillomavirus E2 proteins. Residues in BPV1 E2 that directly contact DNA are indicated with an asterisk. Amino acid substitutions in this region of BPV1 E2 and their effect on DNA binding are shown below.

### I. Transactivation

When joined to a DNA binding domain, the N-terminal 194 amino acids of BPV-1 E2 are able to activate transcription from an E2-responsive promoter. The hinge region of the E2 proteins can be deleted with minimal effects on transactivation. Unlike many transactivation domains, the E2 N-terminal domain seems to have a very constrained structure as almost any deletion that has been made within this domain inactivates all E2-TA functions, presumably by disrupting protein conformation [54,85]. Even certain point mutations (e.g. BPV-1 P106G, G106A) may disrupt protein structure and therefore caution must be used to interpret mutational analyses of the E2 transactivation domain. It is likely, however, that mutations that eliminate one E2 function but not another do not extensively disrupt protein structure.

Recently several groups have undertaken systematic mutational analyses of the transactivation domains of the BPV-1, HPV16 and HPV11 E2 proteins. Several approaches were used and in most cases care was taken to try to avoid mutations that would disrupt protein structure. In one BPV-1 E2 study, conservative substitutions were generated, where possible, for each amino acid that is highly conserved among papillomavirus E2 proteins [12]. A second approach, used for both HPV16 and BPV-1 E2 proteins, was to change conserved charged residues to alanine residues as this can remove an essential side chain with minimal effects on protein structure [1,27,68]. A third study used a yeast screen to select for BPV-1 E2 mutants that were no longer able to activate transcription [11]. And finally, a study of the HPV11 E2 protein used a combination of the first and second approaches and generated both conservative substitutions and alanine residues for each conserved charged residue [84].

## E2 Proteins

The latter study is quite informative as it shows that, in some cases, different substitutions of the same residue can give rise to dissimilar phenotypes. For example, HPV11 E2 proteins R7K and D96E are not defective for transactivation but proteins with alanine at these position have greatly reduced activity. Conversely, substitution of glutamic acid at position 39 with aspartic acid greatly reduces E2 activity but substitution with alanine has no effect. Table 1 shows an abridged summary of mutations in the E2 transactivation domain. This table lists only amino acid residues in which at least one mutation (among the different PV E2 proteins) has a significant effect on transactivation (10% or less of wild type activity). In the studies referenced many more mutations were analyzed and found to have more minimal effects on transactivation. Some proteins were also found to be relatively unstable and the original studies should be consulted for more details. Residues that are highly conserved among the papillomaviruses are shown in bold letters.

As can be seen in figure 1, no short linear sequence in the N-terminal domain seems to be important for transactivation and mutations that eliminate or greatly reduce this function appear to be scattered throughout the domain (with the caveat that some may disrupt overall domain structure). In the study of BPV E2 in which conservative changes were made in highly conserved residues, three mutations (W33F, E39D and K111R) were found to inactivate the transcriptional activation function. (Two other mutations that inactivated E2 function, P106G and G156A, were thought to disrupt protein structure.) However, a W33A mutation in HPV16 E2 gave low but detectable activity suggesting that perhaps the phenylalanine side chain in the BPV E2 mutation interfered with transactivation. Similarly, while aspartic acid or glycine substitutions of residue 39 greatly reduced E2 activity in BPV and HPV11 E2 proteins, all three E2 proteins with alanine residues at this position were not defective. In agreement, however, BPV and HPV11 E2 proteins with K111A or K111R mutations were completely defective. The highly conserved arginine residue at position 37 has also been mutated in all three E2 proteins; R37K was partially defective in HPV11 E2 but not in BPV E2 and R37A resulted in low activity of both BPV1 and HPV11 E2 proteins but no activity in HPV16 E2. The effect of other mutations of the E2 proteins can be seen in table 1. Clearly, more investigation is required to determine whether the differences observed in these studies are due to the different papillomavirus E2 proteins, the different amino acid substitutions or differences in assay conditions. Mutations throughout the transactivation domain were isolated using a yeast screening technique to isolate randomly generated transactivation-defective E2 proteins [11]. This study identified a pattern of bulky hydrophobic (BH) residues in two regions of the E2 protein (residues 26 to 47 and 87 to 107) similar to those that have been found to be important for transcriptional activation in VP16, GAL4 and RTA; mutation of several of these BH residues in E2 inactivated the transcriptional activation function [11]. Other notable mutations in the transactivation domain are a deletion of residues 156 to 159 which eliminates E2 activity but probably disrupts protein conformation; a G156A substitution is also thought to disrupt BPV E2 structure. A four amino acid insertion (PRSR) between residues 181 and 182 of the BPV E2 protein is temperature sensitive; E2 proteins containing this mutation are able to activate transcription at 32° C but not at 39° C [23].

### J. Transcriptional Repression

The BPV1 E2-TR and E8/E2 repressors contain a small portion of the transactivation domain, the hinge region and the C-terminal DNA binding/dimerization domain (see figure 1). Transcriptional repression by these truncated E2 proteins is thought to be due to competitive DNA binding to the E2 binding sites in the viral enhancer elements and to heterodimer formation among transactivator and repressor species (reviewed in [57]. A C-terminal 121 amino acid polypeptide containing the DNA binding/dimerization domain is sufficient for repression of E2-mediated transactivation [52].

A different type of transcriptional repression can be mediated by the full-length transactivator proteins and this depends on the position of E2 binding sites with respect to proximal promoter elements. In many human papillomaviruses two E2 binding sites are positioned between a conserved SP1 site and the TATA box of the major E6/E7 promoter. Binding of E2 to these sites is thought to inhibit binding of these cellular factors resulting in repression of basal promoter activity [6,24,67,77,78]. A C-terminal domain of the E2 proteins is sufficient for this repression in transient assays [24,78].

**Table 1** Transactivation function of mutated E2 proteins

Residue	BPV1	HPV16	HPV11	Reference
R7	A, 35;		A, 10; K, 200;	[27, 84]
Q15	H, 3;			[11]
I30	A, 5;			[27]
L31	P, 5;			[32]
Y32	H, 3;			[32]
<b>W33</b>	F, 0; K, 0;	A, 30;		[12, 27, 68]
<b>R37</b>	K, 140; A, 25;	A, 0;	A, 5; K, 7;	[1, 12, 68, 84]
<b>E39</b>	D, 0; G, 3; A, 65; A, 70;	A, 110;	A, 150; D, 15	[1, 11, 12, 27, 68, 84]
<b>P60</b>	G, 10; A, 30;			[12, 27]
Q66	R, 0;			[32]
<b>I73</b>	L, 20; N, 0; A, 0;	A, 0;		[12, 27, 68]
E74	A, 0; A, 0;		A, 15; D, 5;	[1, 27, 84]
L82	A, 0;			[27]
F87	S, 3;			[11]
E90	A, 45; A, 50;	A, 95;	A, 0; D, 5;	[1, 27, 68, 84]
<b>W92</b>	F, 70; R, 2; A, 0; R, 0;	A, 5;		[11, 12, 27, 68]
S/T93	P, 5;	A, 80;		[32, 68]
D96			A, 5; E, 40;	[84]
W99	C, 3;			[11]
<b>P106</b>	G, 0; S, 0; A, 55;			[12, 27, 32]
<b>K111</b>	R, 0; A, 0;		A, 0; R, 0;	[1, 12, 84]
<b>K112</b>	R, 90; A, 0; A, 0;	A, 20;	R, 40;	[1, 12, 27, 68, 84]
F121	A, 10	A, 80;		[27, 68]
Y131	A, 5;			[27]
Y/W134		A, 10;		[68]
Y138	H, 9;	A, 45;		[11, 68]
W145	R, 1;			[11]
<b>G156</b>	A, 0;			[12]
<b>Y159</b>	F, 110; A, 0;			[12, 27]
156-159(GLYY)		dl, low;		[62]
Y169	A, 0;			[27]
E176/D174	A, 0; A, 50; G, 40;	A, 80;		[1, 27, 32, 68]
S181	F, 10;			[32]
181-182	PRSR insertion, ts;			[23]
R208	G, 10;			[32]

In each column the amino acid substitution is followed by its approximate activity expressed as a % of wild type activity. ts, temperature sensitive; dl, deletion. Highly conserved residues are shown in bold.

## K. Interaction with Cellular Proteins

The full-length BPV1 E2-TA protein has been shown to interact with the cellular replication protein RPA [45] and the cellular transcription factor Sp1 [48]. The C-terminal 127 amino acids of E2 or fusion proteins containing the C-terminal 160 amino acids of E2 are unable to interact with Sp1, implying that the N-terminal domain is required for this interaction [48].

## E2 Proteins

The BPV1 E2 proteins interact with and cooperatively bind to DNA with the cellular basal transcription factors, TFIID and TFIIB [65,74]. The DNA binding domain of E2 (residues 310–410) is sufficient for protein-protein interaction with these factors [65]. However, the E2 hinge region is required in addition to the DNA binding domain (residues 204–410) for cooperative binding to DNA with TFIID or TFIIB [65,74].

### L. Replication

Almost all studies of papillomavirus replication have examined E2 functions that are important for the plasmid maintenance function of the virus rather than vegetative viral DNA replication. However, using a cell culture system in which viral genomes are amplified in a subpopulation of cells, Alderborn *et al.* demonstrated that a temperature sensitive BPV1 E2 protein (see tables 1 and 2) that is defective for transactivation and plasmid maintenance replication at the non-permissive temperature was able to amplify large amounts of the viral genome in division arrested cells [2].

Most other studies have examined the ability of E2 to replicate plasmids in a transiently transfected cells. The full-length E2 protein is necessary for viral DNA replication [81]. The E2 protein binds cooperatively to the replication origin with the E1 protein [51,59,72,87], interacts with at least one cellular replication protein, RPA [45] and alleviates nucleosomal-mediated repression of replication [46]. The HPV E2 proteins seem to play a similar role in replication and, in fact, certain combinations of BPV1 and HPV E1 and E2 proteins are capable of initiating replication from various papillomavirus origins [16,19,30]. This indicates that the replication functions of these proteins are quite well conserved. Efficient replication by the BPV1 proteins *in vivo* requires the E2 protein. However, E2 is not necessary for replication of naked DNA templates *in vitro*, although it can enhance replication at low concentrations of the E1 protein [87,88]. In addition, it has been shown that the E1 protein of HPV1a is sufficient for replication *in vivo* [30]. Therefore, it appears that E2 plays an auxiliary role in replication and that E1 is the principal replication protein.

The transactivation domain of E2 is absolutely required for DNA replication [85]. In several studies, amino acid substitutions have been generated in the transactivation domains of the BPV1, HPV11 and HPV16 E2 proteins to determine which regions of this domain are important for the replication function(s) [1,11,12,27,32,68,84]. In the interest of space, only those mutated proteins that have 10% or less of wild type replication activity in at least one papillomavirus E2 protein are represented here. As shown in table 2, mutations throughout the domain affect the replication properties of the E2 protein but mutations in the same residue do not always give similar phenotypes in the different E2 proteins. Clearly, more investigation is required to determine whether the differences observed in these studies are due to the different papillomavirus E2 proteins, the different amino acid substitutions or differences in assay conditions. In many cases proteins defective for replication are also defective for transactivation but this is not always the case and separation of these properties are discussed in section N below. Two of the studies [12,68], also examined the ability of these mutated E2 proteins to interact with the E1 protein. In the BPV1 study, proteins that were defective for replication could still interact with the E1 protein (W33F, E39D, K111R). However, in the HPV16 study many of the replication-defective proteins with alanine substitutions could not interact efficiently with the E1 protein.

No particular amino acid sequence of the hinge region of the E2 protein is required for DNA replication. However, some nonspecific sequence is required between the two conserved domains to maintain the replication function. Two proteins with large deletions of the hinge region ( $E2_{\Delta 220-309}$  and  $E2_{\Delta 213-309}$ ) are unable to promote replication yet they can activate transcription more efficiently than the wildtype E2 protein [85]. It is possible that the two domains are too closely linked which sterically hinders one or more replication function.

In most cases, an intact E2 DNA binding domain is required for BPV1 DNA replication [82,85] and an E2 DNA binding site is required in the replication origin [80]. However, several E2 proteins have been identified that are defective in DNA binding but can support DNA replication to some extent. In one study, two out of ten E2 proteins with deletions in the DNA binding domain were able to support DNA replication at low levels [85]. In another study, DNA binding defective E2 proteins with point mutations in the DNA binding domain were able to promote replication *in vitro* [49].

**Table 2** Replication function of mutated E2 proteins

Residue	BPV1	HPV16	HPV11	Reference
Q15	H, very low;			[32]
I30	A, 10;			[27]
<b>W33</b>	F, 5; K, 3;	A, 20;		[12, 27, 68]
<b>E39</b>	D, 5; G, 100; A, 125; A;15;	A, 10;	A, 15; D, 10;	[1, 12, 27, 32, 68, 84]
Q66	R, 5;			[32]
L82	A, 5;			[27]
<b>W92</b>	F, 45; R, 0; R, 5; A, 5;	A, 20;		[12, 27, 32, 68]
W99	C, 5;			[32]
<b>P106</b>	G, 0; S, 0; A, 100;			[12, 27, 32]
<b>K111</b>	R, 5; A, 0;		A, 5; R, 5;	[1, 12, 84]
<b>K112</b>	R, 60; A, 0; A, 5;	A, 65;	R, 100;	[1, 12, 27, 68, 84]
F121	A, 5;	A, 30;		[27, 68]
Y131	A, 10;			[27]
W145	R, 5;			[32]
<b>G156</b>	A, 0;			[12]
Y169	A, 0;			[27]
E/D176	G, 100; A, 0; A, 60;	A, 95; (D174)		[1, 27, 32, 68]
181-182	PRSR insertion, ts;			[23]

In each column the amino acid substitution is followed by its activity expressed as a % of wild type activity. ts, temperature sensitive. Highly conserved residues are shown in bold.

BPV1 E2 proteins with mutations in the redox sensitive cysteine residue at position 340 in the DNA binding domain are able to support replication but are unable to activate transcription [31]. However, it is likely that the DNA binding and dimerization properties of the E2 protein are important for its replication function in the complete viral life cycle; the position of the E2 binding sites with respect to the E1 binding site is conserved among papillomavirus origins which indicates that they have an important function. It is possible that the mutated proteins have acquired some property that allows them to compensate for the absence of DNA binding, such as increased stability or increased interaction with the E1 protein. In addition, these experiments have been carried out using *in vitro* or transient assays in which the E1 and E2 proteins are most likely expressed at quite high levels and the replicon DNA is probably not assembled completely in chromatin and therefore all functions of the E2 proteins may not be required.

It is not clear whether the E2 repressors can repress replication. Disruption of E2-TR expression in BPV1 results in a virus that replicates at much higher copy number than wild type [40,66] however, this effect could be indirect and due to lack of transcriptional repression. Notably, the E2 sites flanking the origin in BPV1 (sites 11 and 12) have a relatively weak affinity for the E2 protein. Binding of E2 to these sites is greatly increased in the presence of the E1 protein. This may be important for viral replication to ensure that only the E1-E2 complex, and not the E2-TR repressor protein, can bind to the origin region with high affinity.

#### M. Interaction and cooperative binding to the replication origin with the E1 protein.

The viral DNA replication origin contains an E1 binding site flanked by E2 binding sites [82]. The E1 and E2 proteins interact to form a protein complex and bind cooperatively to the origin of replication [7,51,59,72,87]. Formation of the replication preinitiation complex requires specific protein-protein and

## E2 Proteins

protein-DNA interactions between the E1 and E2 proteins and their respective DNA binding motifs. A biochemical study found that the entire transactivation domain and the DNA binding domain of BPV1 E2 are required for enhancement of E1 binding to the origin region (the hinge is not necessary). The entire transactivation domain is necessary and probably sufficient for interaction with E1 and the DNA binding domain is required for binding the complex to the origin DNA [86]. The E2-TR protein is unable to interact with or cooperatively bind to the origin with the E1 protein [59,86]. Another biochemical study showed that the N-terminal 140 amino acids of HPV16 E2 were unable to interact with E1 [75]. However, the first 91 amino acids of BPV1 E2 were able to interact with E1 in the yeast two hybrid system [5]. Two classes of mutations in the transactivation domain have been identified that interfere with the E1-E2 interaction. Mutations in the first class may identify regions of the E1 protein that are involved in protein-protein interaction because, although they are defective in E1 binding, they are transactivation competent [12,68]. These mutations identify a region of E2 in the vicinity of residues 20, 33 and 39 and a second region containing residues 178 and 188 as being important for E1 interaction (see table 3). Notably, an antibody against HPV16 E2 residues 18 to 41 is able to block the E1-E2 interaction, supporting the fact that this region is important for complex formation [35]. The second class of E2 mutations that eliminate E1-E2 interaction are more difficult to interpret (see Table 3). These mutations are also defective in transcriptional activation and/or replication and the overall conformation of the N-terminal domain may be disrupted. Notably, these mutations are either deletions or mutations of or to prolines or glycines.

**Table 3 E1 binding properties of mutated E2 proteins**

Mutated E2 proteins that are defective in E1 interaction but functional in other E2 properties	Reference	Mutated E2 proteins that are defective both in E1 interaction and in other E2 properties	Reference
BPV1 E20D	[12]	HPV16 Δ23–26 HPV16	[62]
HPV16 W33A	[68]	L26P	[62]
HPV16 E39A	[68]	BPV1 P106G	[12]
HPV16 Y178A	[68]	BPV1 G106A	[12]
BPV1 V188L	[12]	HPV16 156–159	[62, 75]

## N. Separation of transactivation and replication properties

The E2-TA protein is important for transcriptional transactivation and DNA replication and several E2 mutants have been isolated that separate these properties. E2 proteins with deletions of the entire hinge region are able to activate transcription at wildtype levels yet are unable to support replication [85]. Conversely, a subset of proteins with deletions in the DNA binding domain of E2 are unable to activate transcription, yet can support DNA replication [85]. Other mutated E2 proteins with amino acid substitutions in the DNA binding domain (R344L, C340F) are also unable to activate transcription but can enhance *in vitro* replication [49]. Proteins with mutations in the redox-sensitive C340 are defective in transactivation in mammalian cells yet they can support viral DNA replication [31]. Several point mutations in the N-terminal domain also separate the transactivation and replication properties of E2 and these are listed in Table 4. Only mutations that are quite defective in one function (5% or less activity) and show reasonably high levels of activity in the other function are represented here. The cited studies should be referred to for additional mutated proteins that are low in one activity and high in the others. This summary shows that the requirements for replication seem to be much less stringent than those for activation of transcription. Two regions appear to be important for transactivation but not replication; mutations in residues R37, I73 and/or E74 separate these functions in all three E2 proteins (BPV-1 E2 R37A also shows differential activity but as it retains 20% transactivation activity it is not

shown in Table 4). Only two categories of mutants have been identified that cannot support DNA replication but can activate transcription. Proteins in one category have deletions of the entire hinge region and it has been postulated that this may cause some steric hindrance that interferes with one or more of the functions required for replication [85]. An HPV16 E2 protein with an E39A substitution is also defective for DNA replication and not transactivation but this is probably due to the inability of this protein to interact with the E1 protein [68]. (The same mutation in BPV-1 and HPV11 E2 proteins also reduces replication activity to approximately 15% but does not greatly affect transactivation.) These results probably reflect the fact that E2 plays a primary role in transcriptional activation but only an auxiliary role in DNA replication.

---



---

**Table 4 Separation of transactivation and replication properties of the E2 proteins**

E2 proteins that can support replication but not transactivation	Reference	E2 proteins that can activate transcription but not support DNA replication	Reference
<b>BPV1</b>		<b>BPV1</b>	
L31P, Y32H, E39G, F87S, S93P, E74A	[32] [1]	E2 <sub>Δ220–309</sub> , E2 <sub>Δ212–309</sub>	[85]
I73N, I73A, E74A, E2 <sub>1–210</sub> , E2 <sub>1–376</sub> C340S, C340G	[27] [85] [31]	<b>HPV16</b> E39A	[68]
<b>HPV16</b>			
R37A, I73A, W92A	[68]		
<b>HPV11</b>			
R37A, E74D, E90A, E90D, D96A	[84]		

Only those mutated proteins that are almost completely inactive in one assay (5% or less than wild type activity) are shown.

---



---

## O. Growth suppression by the papillomavirus E2 proteins

The BPV1 E2-TA protein is able to suppress the growth of HeLa cells, a line that is derived from an HPV-containing cervical carcinoma [26,37,79]. HeLa cells are dependent on expression of the endogenous HPV18 E6 and E7 proteins for continued cell growth and it is thought likely that E2 suppresses growth, at least in part, by repressing transcription of the E6/E7 P105 promoter. An intact transactivation and DNA binding domain are required for growth suppression even though the E2-TR repressor is able to repress expression from the P105 promoter in transient assays [26]. One explanation for this difference is that the E2 transactivation domain may be required to alleviate nucleosomal repression of the integrated HPV18 genome [26]. E2 may also inhibit cell growth by other mechanisms as one study has found that E2 expressed from a recombinant virus can also inhibit growth of an HPV-negative cervical carcinoma line [37]. This was not found when E2-containing HPV-negative cells were isolated by drug selection, perhaps because of the different E2 expression levels [26].

## P. Alleviation of Nucleosomal Repression

The E2-TA protein can antagonize nucleosomal repression of BPV1 DNA replication *in vitro* [46]. This function depends on the presence of E2 DNA binding sites in the origin and therefore probably requires the E2 DNA binding domain. It is not known whether the E2 transactivation domain is required for this function but in general the transactivation functions of cellular factors are required to relieve nucleosomal repression.

## E2 Proteins

### Q. Cooperative DNA binding

BPV1 E2-TA binds cooperatively to two adjacent DNA binding sites with a cooperativity parameter of 8.5. The 86 amino acid DNA binding domain and the E2-TR protein exhibit much less cooperativity (factors of 1.9 and 2.9, respectively) which implies that the N-terminal domain of E2 is important for this function [60]. However, this cooperativity of DNA binding is not sufficient to explain the great synergy of transcription obtained with one versus two DNA binding sites.

### R. Looping

The BPV1 E2-TA protein can form stable loops between widely spaced DNA sites that are visible by electron microscopy [38]. The shorter E2-TR and E8-E2 proteins are unable to form such loops implying that the transactivation domain is required for this function.

### S. B and T cell epitopes

A number of linear B cell epitopes have been mapped in the HPV16 E2 protein. The E2 open reading frame (ORF) was synthesized as a set of overlapping 20-residue peptides which were tested for reactivity with HPV16-infected patient sera [20]. The E2 ORF is the most reactive of all the HPV16 ORFs and four of the most reactive peptides (E2:9, residues 121–140; E2:13, residues 181–200; E2:17, residues 241–260; E2:19 residues 271–290) are shown in figure 1. In some instances, specific epitopes are recognized preferentially by either IgG, IgA or IgM antibodies. Figure 1 also shows another major IgG and IgA reactive epitope from HPV16 and HPV18 E2, designated p245 (residues 328–345) [21]. There is an association between serum antibodies against some of these epitopes and HPV-associated lesions and carcinomas [22,34,42,44,76,83]. A T-helper cell epitope overlaps the p245 B-cell epitope in HPV18 E2 but not in HPV16 E2 [43]. As summarized in figure 1, four T-helper cell epitopes have also been identified in HPV16 E2 (residues 11–25, 141–155, 191–205 and 231–245) and shown to overlap with additional IgG specific B-cell epitopes [22].

### References

- [1] Abroi, A., Kurg, R., Ustav, M. (1996) Transcriptional and replicational activation functions in the BPV1 E2 protein are encoded by different structural determinants. *J Virol* **70**:6169.
- [2] Alderborn, A., Jareborg, N., Burnett, S. (1992) Evidence that the transcriptional trans-activating function of the bovine papillomavirus type 1 E2 gene is not required for viral DNA amplification in division-arrested cells. *J Gen Virol* **73**:2639–51.
- [3] Baker C., Calef C. (1995) Maps of papillomavirus mRNA transcripts. In Myers G., Bernard H-U., Baker C., Halpern A., Delius H., Icenogel J., editors, *Human Papillomaviruses 1995*. Los Alamos National Laboratory, LA-UR 95-3675 pp. III-3–III-20.
- [4] Barbosa, M.S., Wettstein, F.O. (1988) E2 of cottontail rabbit papillomavirus is a nuclear phosphoprotein translated from an mRNA encoding multiple open reading frames. *J Virol* **62**:3242–9.
- [5] Benson, J.D., Howley, P.M. (1995) Amino-terminal domains of the bovine papillomavirus type 1 E1 and E2 proteins participate in complex formation. *J Virol* **69**:4364–72.
- [6] Bernard, B.A., Bailly, C., Lenoir, M.-C., Darmon, M., Thierry, F., Yaniv, M. (1989) The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *J Virol* **63**:4317–24.
- [7] Blitz, I.L., Laimins, L.A. (1991) The 68-kilodalton E1 protein of bovine papillomavirus is a DNA binding phosphoprotein which associates with the E2 transcriptional activator in vitro. *J Virol* **65**:649–56.
- [8] Bochkarev, A., Barwell, J.A., Pfuetzner, R.A., Furey, W.J., Edwards, A.M., Frappier, L. (1995) Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA 1. *Cell* **83**:39–46.

- [9] Bouvard, V., Storey, A., Pim, D., Banks, L. (1995) Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *EMBO J* **13**:5451–9.
- [10] Bream, G.L., Ohmstede, C.A., Phelps, W.C. (1993) Characterization of human papillomavirus type 11 E1 and E2 proteins expressed in insect cells. *J Virol* **67**:2655–63.
- [11] Breiding, D.E., Grossel, M.J., Androphy, E.J. (1996) Genetic analysis of the bovine papillomavirus E2 transcriptional activation domain. *Virology* **221**:34–43.
- [12] Brokaw, J.L., Blanco, M., McBride, A.A. (1996) Amino acids critical for the functions of the BPV-1 E2 transactivator. *Journal of Virology* **70**:23–9.
- [13] Burnett, S., Ström, A.-C., Jareborg, N., Alderborn, A., Dillner, J., Moreno-Lopez, J., Pettersson, U., Kiessling, U. (1990) Induction of bovine papillomavirus E2 gene expression and early region transcription by cell growth arrest: Correlation with viral DNA amplification and evidence for differential promoter induction. *J Virol* **64**:5529–41.
- [14] Carruth, M., McBride, A.A. (1996) unpublished data
- [15] Chiang, C.-M., Broker, T.R., Chow, L.T. (1991) An E1M^E2C fusion protein encoded by human papillomavirus type 11 is a sequence-specific transcription repressor. *J Virol* **65**:3317–29.
- [16] Chiang, C.M., Ustav, M., Stenlund, A., Ho, T.F., Broker, T.R., Chow, L.T. (1992) Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins. *Proc Natl Acad Sci USA* **89**:5799–803.
- [17] Choe, J., Vaillancourt, P., Stenlund, A., Botchan, M. (1989) Bovine papillomavirus type 1 encodes two forms of a transcriptional repressor: Structural and functional analysis of new viral cDNAs. *J Virol* **63**:1743–55.
- [18] Corina, K., Grossman, S.R., Barsoum, J., Prakash, S.S., Androphy, E.J., Pepinsky, R.B. (1993) The tryptophan bridge is a critical feature of the papillomavirus E2 DNA binding domain. *Virology* **197**:391–6.
- [19] Del Vecchio, A.M., Romanczuk, H., Howley, P.M., Baker, C.C. (1992) Transient replication of human papillomavirus DNAs. *J Virol* **66**:5949–58.
- [20] Dillner, J. (1990) Mapping of linear epitopes of human papillomavirus type 16: The E1, E2, E4, E5, E6 and E7 open reading frames. *Int J Cancer* **46**:703–11.
- [21] Dillner, J., Dillner, L., Robb, J., Willems, J., Jones, I., Lancaster, W., Smith, R., Lerner, R. (1989) A synthetic peptide defines a serologic IgA response to a human papillomavirus-encoded nuclear antigen expressed in virus-carrying cervical neoplasia. *Proc Natl Acad Sci USA* **86**:3838–41.
- [22] Dillner, L., Zellbi, A., Avall-Lundqvist, E., Heino, P., Eklund, C., Pettersson, C.A., Forslund, O., Hansson, B.G., Grandien, M., Bistoletti, P. (1995) Association of serum antibodies against defined epitopes of human papillomavirus L1, E2, and E7 antigens and of HPV DNA with incident cervical cancer. *Cancer Detect Prev* **19**:381–93.
- [23] DiMaio, D., Settleman, J. (1988) Bovine papillomavirus mutant temperature sensitive for transformation, replication and transactivation. *EMBO J* **7**:1197–204.
- [24] Dong, G., Broker, T.R., Chow, L.T. (1995) Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J Virol* 1994 Feb **68**:1115–27.
- [25] Doorbar, J., Parton, A., Hartley, K., Banks, L., Crook, T., Stanley, M., Crawford, L. (1990) Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology* **178**:254–62.
- [26] Dowhanick, J.J., McBride, A.A., Howley, P.M. (1995) Suppression of cellular proliferation by the papillomavirus E2 protein. *J Virol* **69**:7791–9.
- [27] Ferguson, M.F., Botchan, M.R. (1996) Genetic analysis of the activation domain of bovine papillomavirus protein E2:its role in transcription and replication. *J Virol* **70**:4193–9.
- [28] Gauthier, J.M., Dillner, J., Yaniv, M. (1991) Structural analysis of the human papillomavirus type 16-E2 transactivator with antipeptide antibodies reveals a high mobility region linking the transactivation and the DNA-binding domains. *Nucleic Acids Res* **19**:7073–9.

## E2 Proteins

- [29] Giri, I., Yaniv, M. (1988) Structural and mutational analysis of E2 trans-activating proteins of papillomaviruses reveals three distinct functional domains. *EMBO J* **7**:2823–9.
- [30] Gopalakrishnan, V., Khan, S.A. (1995) E1 protein of human papillomavirus type 1a is sufficient for initiation of viral DNA replication. *Proc Natl Acad Sci USA* **91**:9597–601.
- [31] Grossel, M.J., Barsoum, J., Prakash, S.S., Androphy, E.J. (1996) The BPV-1 E2 DNA-contact helix cysteine is required for transcriptional activation but not replication in mammalian cells. *Virology* **217**:301–10.
- [32] Grossel MJ, Sverdrup F, Breiding DE, Androphy EJ. Transcriptional activation function is not required for stimulation of DNA replication by BPV-1 E2. *J Virol* (In Press) 1996;
- [33] Hegde, R.S., Grossman, S.R., Laimins, L.A., Sigler, P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target [see comments]. *Nature* **359**:505–12.
- [34] Heino, P., Eklund, C., Fredriksson-Shanazarian, V., Goldman, S., Schiller, J.T., Dillner, J. (1995) Association of serum immunoglobulin G antibodies against human papillomavirus type 16 capsids with anal epidermoid carcinoma. *J Natl Cancer Inst* **87**:437–40.
- [35] Hibma, M.J., Raj, K., JEly, S.Y., Stanley, J., Crawford, L. (1995) The interaction between human papillomavirus type 16 E1 and E2 proteins is blocked by an antibody to the N-terminal region of E2. *Eur J Biochem* **229**:517–25.
- [36] Hubbert, N.L., Schiller, J.T., Lowy, D.R., Androphy, E.J. (1988) Bovine papilloma virus-transformed cells contain multiple E2 proteins. *Proc Natl Acad Sci USA* **85**:5864–8.
- [37] Hwang, E.S., Riese, D.J., Settleman, J., Nilson, L.A., Honig, J., Flynn, S., DiMaio, D. (1993) Inhibition of cervical carcinoma cell line proliferation by the introduction of a bovine papillomavirus regulatory gene. *J Virol* **67**:3720–9.
- [38] Knight, J.D., Li, R., Botchan, M. (1991) The activation domain of the bovine papillomavirus E2 protein mediates association of DNA-bound dimers to form DNA loops. *Proc Natl Acad Sci USA* **88**:3204–8.
- [39] Lambert, P.F., Hubbert, N.L., Howley, P.M., Schiller, J.T. (1989) Genetic assignment of multiple E2 gene products in bovine papillomavirus-transformed cells. *J Virol* **63**:3151–4.
- [40] Lambert, P.F., Monk, B.C., Howley, P.M. (1990) Phenotypic analysis of bovine papillomavirus type 1 E2 repressor mutants. *J Virol* **64**:950–6.
- [41] Lambert, P.F., Spalholz, B.A., Howley, P.M. (1987) A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator. *Cell* **50**:69–78.
- [42] Lehtinen, M., Leminen, A., Paavonen, J., Lehtovirta, P., Hyoty, H., Vesterinen, E., Dillner, J. (1995) Predominance of serum antibodies to synthetic peptide stemming from HPV 18 open reading frame E2 in cervical adenocarcinoma. *J Clin Pathol* 1992 Jun **45**:494–7.
- [43] Lehtinen, M., Stellato, G., Hyoty, H., Nieminen, P., Vesterinen, E., Paavonen, J. (1992) A T-helper cell epitope overlaps a major B-cell epitope in human papillomavirus type 18 E2 protein. *APMIS* **100**:1022–6.
- [44] Lenner, P., Dillner, J., Wiklund, F., Hallmans, G., Stendahl, U. (1995) Serum antibody responses against human papillomavirus in relation to tumor characteristics, response to treatment, and survival in carcinoma of the uterine cervix. *Cancer Immunol Immunother* **40**:201–5.
- [45] Li, R., Botchan, M.R. (1993) The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication. *Cell* **73**:1207–21.
- [46] Li, R., Botchan, M.R. (1994) Acidic transcription factors alleviate nucleosome-mediated repression of DNA replication of bovine papillomavirus type 1. *Proc Natl Acad Sci USA* **91**:7051–5.
- [47] Li, R., Knight, J., Bream, G., Stenlund, A., Botchan, M. (1989) Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev* **3**:510–26.

- [48] Li, R., Knight, J.D., Jackson, S.P., Tjian, R., Botchan, M.R. (1991) Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. *Cell* **65**:493–505.
- [49] Li, R., Yang, L., Fouts, E., Botchan, M.R. (1993) Site-specific DNA-binding proteins important for replication and transcription have multiple activities. *Cold Spring Harb Symp Quant Biol* **58**:403–13.
- [50] Liang, H., Petros, A.M., Meadows, R.P., Yoon, H.S., Egan, D.A., Walter, K., Holzman, T.F., Robins, T., Fesik, S.W. (1996) Solution structure of the DNA-binding domain of a human papillomavirus E2 protein –evidence for flexible DNA-binding regions. *Biochemistry* **35**:2095–103.
- [51] Lusky, M., Hurwitz, J., Seo, Y. (1993) Cooperative assembly of the bovine papilloma virus E1 and E2 proteins on the replication origin requires an intact E2 binding site. *J Biol Chem* **268**:15795–803.
- [52] McBride, A.A. (1996) unpublished observations
- [53] McBride, A.A., Bolen, J.B., Howley, P.M. (1989) Phosphorylation sites of the E2 transcriptional regulatory proteins of bovine papillomavirus type 1. *J Virol* **63**:5076–85.
- [54] McBride, A.A., Byrne, J.C., Howley, P.M. (1989) E2 polypeptides encoded by bovine papillomavirus type 1 form dimers through the common carboxyl-terminal domain: Transactivation is mediated by the conserved amino-terminal domain. *Proc Natl Acad Sci USA* **86**:510–4.
- [55] McBride, A.A., Howley, P.M. (1991) Bovine papillomavirus with a mutation in the E2 serine 301 phosphorylation site replicates at a high copy number. *J Virol* **65**:6528–34.
- [56] McBride, A.A., Klausner, R.D., Howley, P.M. (1992) Conserved cysteine residue in the DNA-binding domain of the bovine papillomavirus type 1 E2 protein confers redox regulation of the DNA-binding activity in vitro. *Proc Natl Acad Sci USA* **89**:7531–5.
- [57] McBride, A.A., Romanczuk, H.R., Howley, P.M. (1991) The Papillomavirus E2 Regulatory Proteins. *J Biol Chem* **266**:18411–4.
- [58] Meneguzzi, G., Lathe, R., Kiely, M.P., Vogt, N. (1989) The E2 trans-activating protein of bovine papillomavirus type 1 (BPV1) is serine-phosphorylated in vivo. *Oncogene* **4**:1285–90.
- [59] Mohr, I.J., Clark, R., Sun, S., Androphy, E.J., MacPherson, P., Botchan, M.R. (1990) Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* **250**:1694–9.
- [60] Monini, P., Grossman, S.R., Pepinsky, B., Androphy, E.J., Laimins, L.A. (1991) Cooperative binding of the E2 protein of bovine papillomavirus to adjacent E2-responsive sequences. *J Virol* **65**:2124–30.
- [54] O'Connor M, Chan S, Bernard H. (1995) *Transcription factor binding sites in the long control region of the genital HPVs.* in Myers G, Bernard H, Baker C, Halpern A, Delius H, Icenogel J, editors. *Human Papillomaviruses 1995.* Los Alamos National Laboratory, LA-UR 95-3675, pp. III-21–40.
- [62] Piccini, A., Storey, A., Massima, P., Banks, L. (1995) Mutations in the human papillomavirus type 16 E2 protein identify multiple regions of the protein involved in binding to E1. *J Gen Virol* **76**:2909–13.
- [63] Piirsoo, M., Ustav, E., Mandel, T., Stenlund, A., Ustav, M. (1996) Cis and trans requirements for stable episomal maintenance of the BPV-1 replicator. *EMBO J* **15**:1–11.
- [64] Prakash, S.S., Grossman, S.R., Pepinsky, R.B., Laimins, L.A., Androphy, E.J. (1992) Amino acids necessary for DNA contact and dimerization imply novel motifs in the papillomavirus E2 transactivator. *Genes Dev* **6**:105–16.
- [65] Rank, N.M., Lambert, P.F. (1995) Bovine papillomavirus type 1 E2 transcriptional regulators directly bind two cellular transcription factors, TFIID and TFIIB. *J Virol* **69**:6323–34.
- [66] Riese, D.J., II, Settleman, J., Neary, K., DiMaio, D. (1990) Bovine papillomavirus E2 repressor mutant displays a high-copy-number phenotype and enhanced transforming activity. *J Virol* **64**:944–9.

## E2 Proteins

- [67] Romanczuk, H., Thierry, F., Howley, P.M. (1990) Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J Virol* **64**:2849–59.
- [68] Sakai, H., Yasugi, T., Benson, J.D., Dowhanick, J.J., Howley, P.M. (1996) Targeted mutagenesis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions. *J Virol* **70**:1602–11.
- [69] Sanders, C.M., Stern, P.L., Maitland, N.J. (1995) Characterization of human papillomavirus type 16 E2 protein and subdomains expressed in insect cells. *Virology* **211**:418–33.
- [70] Schiller, J. Personal Communication.
- [71] Skiadopoulos, M.H., McBride, A.A. (1996) The BPV-1 E2 transactivator and E2-TR repressor use different nuclear localization signals. *J Virol* **70**:1117–24.
- [72] Spalholz, B.A., McBride, A.A., Sarafi, T., Quintero, J. (1993) Binding of bovine papillomavirus E1 to the origin is not sufficient for DNA replication. *Virology* **193**:201–12.
- [73] Spalholz, B.A., Yang, Y.C., Howley, P.M. (1985) Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. *Cell* **42**:183–91.
- [74] Steger, G., Ham, J., Lefebvre, O., Yaniv, M. (1995) The bovine papillomavirus 1 E2 protein contains two activation domains: one that interacts with TBP and another that functions after TBP binding. *EMBO J* **14**:329–40.
- [75] Storey, A., Piccini, A., Massima, P., Bouvard, V., Banks, L. (1995) Mutations in the human papillomavirus type 16 E2 protein identify a region of the protein involved in binding to E1 protein. *J Gen Virol* **76**:819–26.
- [76] Strickler, H.D., Dillner, J., Schiffman, M.H., Eklund, C., Glass, A.G., Greer, C., Scott, D.R., Sherman, M.E., Kurman, R.J., Manos, M. (1994) A seroepidemiologic study of HPV infection and incident cervical squamous intraepithelial lesions. *Viral Immunol* **7**:169–77.
- [77] Tan, S.H., Leong, L.E., Walker, P.A., Bernard, H.U. (1994) The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J Virol* **68**:6411–20.
- [78] Thierry, F., Howley, P.M. (1991) Functional analysis of E2-mediated repression of the HPV18 P105 promoter. *New Biol* **3**:90–100.
- [79] Thierry, F., Yaniv, M. (1987) The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. *EMBO J* **6**:3391–7.
- [80] Ustav, E., Ustav, M., Szymanski, P., Stenlund, A. (1993) The bovine papillomavirus origin of replication requires a binding site for the E2 transcriptional activator. *Proc Natl Acad Sci USA* **90**:898–902.
- [81] Ustav, M., Stenlund, A. (1991) Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. *EMBO J* **10**:449–57.
- [82] Ustav, M., Ustav, E., Szymanski, P., Stenlund, A. (1991) Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. *EMBO J* **10**:4321–9.
- [83] Wikstrom, A., Eklund, C., von Krogh, G., Lidbrink, P., Dillner, J. (1992) Levels of immunoglobulin G antibodies against defined epitopes of the L1 and L2 capsid proteins of human papillomavirus type 6 are elevated in men with a history of condylomata acuminata. *J Clin Microbiol* **30**:1795–800.
- [84] Winokur, P. Personal Communication.
- [85] Winokur, P.L., McBride, A.A. (1992) Separation of the transcriptional activation and replication functions of the bovine papillomavirus-1 E2 protein. *EMBO J* **11**:4111–8.
- [86] Winokur, P.L., McBride, A.A. (1996) The transactivation and DNA binding domains of the BPV-1 E2 protein have different roles in co-operative origin binding with the E1 protein. *Virology* **221**:44–53.

- [87] Yang, L., Li, R., Mohr, I.J., Clark, R., Botchan, M.R. (1991) Activation of BPV-1 replication in vitro by the transcription factor E2. *Nature* **353**:628–32.
- [88] Yang, L., Mohr, I., Fouts, E., Lim, D.A., Nohaile, M., Botchan, M. (1995) The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase. *Proc Natl Acad Sci USA* **90**:5086–90.
- [89] Yang, L., Mohr, I., Li, R., Nottoli, T., Sun, S., Botchan, M. (1991) Transcription factor E2 regulates BPV-1 DNA replication in vitro by direct protein-protein interaction. *Cold Spring Harb Symp Quant Biol* LVI:335–46.
- [90] Zhou, J. Personal Communication.

## **E2 Appendix A**

### **Appendix A: E2 Amino Acid Sequence Alignment**

The following alignment of E2 amino acid sequences was generated using a Hidden Markov Method (HMM) of analysis, as described in another section of Part III (Farmer and Myers). The alignment differs from what is provided in Part II insofar as A and B supergroup sequences have been aligned together by the algorithm (whereas in Part II they are separately aligned). This alignment should be viewed as merely an hypothesis; furthermore, it becomes the basis for the structural prediction output in the following appendix (appendix B).

MOST-LIKELY M.....	ETLSERLDALQEKLLELYE.KDSKLEDQIEHWKLLRLENVLLYKAREMGITRLGHQVVPLAVSK	66
HPV54	-.....-AT--VC--R--D--.-..-NK-----CI---CA-Q-----YKV-Q--AL-A-----	66
HPV32	-.....-AK---C-Q----.E---H--KHVQ---C---I-AA--F-----YAQV---I--A-EI-R	66
HPV42	-.....-R-AK---C-Q----.EN-R--QKH----C---M-A-V-----FANI---I--T-ETCR	66
HPV3	-.....-AN--VC-D-I----.DK----M--Q-M--QA-----C-L-HI-----S-T-	66
HPV28	-.....-AN--VC-D-M----.NK----M--Q-M-V--A-----C-L-HI-----S-T-	66
HPV10	-.....-AN--C-D-M----.DK----T--H--V--A-----C-L-HI-----S-T-	66
HPV29	-.....-N-AN---C-D-I----.R--DK----T--Y-M-V-SA-Y-----C-M--I-----T-S-A-	66
HPV61	-R....M-S-AD--C---D--.-..-NK----L--HYV---AM-F--QA-L--V--M--T-S-T-	68
HPV2a	-.....-AN---C-T----.NK----K--AQV----M-F----C-M--V-CTA--A-T---	66
HPV27	-.....-AN---C-T----.NK----K--AQV----M-F----C-M--V-CTT--A-T---	66
HPV57	-.....-AS---C-T----.NK----K--AQV----M-F----C-M--V-CTT--A-T---	66
HPV26	-.....-N-CQ--N-C--I-DY--.L--NK-T--DY--V-Y-CAIF--GNMQCIN----STV-C	66
HPV51	-.....-CH--NVC---I-DC--.L--DK-V--NY-T--Y-AAMP-A--RNLRITN----ATT--	66
HPV30	-.....-CQ---C---I-DCF--.N--KI--H-VY--AV-H--V-----QNN--K-R-----C-Q-C-	66
HPV53	-.....-CQ---C---I-DCF--.R--NIT-H-DY--AV-Q--IY--NNM-K-----C-Q-C-	66
HPV56	-.....-Q--N-C-N-I-DCF-X--RCIA-H--Y--AV-H--Y-----ND--V--N--M--C-Q-C-	67
HPV66	-.....-Q---C-N-I-DC--.---CII-H-DY--AV-H-Y--Y-----ND-NV-N--M--S-Q-C-	66
HPV18	-QT..BK---S--D-I-IDH--.N--IDS--QY-Q-I-W--AIFFA--H--QT-N-----AYNI--	70
HPV45	-KMQTPK-S--S--D-I-DH--.N----INS--SY-Q-I--AI-FT--H--K-N-----INI--	72
HPV39	-KE.TMMK--Q--NV--D-I--Y--.Q--SIY--NY--CV-M--AIF-A--R-MHTID----TINI--	71
HPV70	-KE.TMM--Q--N--I--H--.Q--LIY--NY--YV--AIF-A--R-MHTID----GTT--	71
HPV59	-QT..VMD--Q--SV--DQI--H--.N--INEH-NY--V-M--I-FA--NN-HT-N-----TFL--	70
HPV7	-.....-K-AR--LC--Q----.Q--Q-QHH--L--YI-Y-S--IY-T--Q--K-----S-D--	66
HPV40	-.....-N-AR--LC--Q----.QN--E-QQH--L--YI-Y-SAIY-T--Q--K--K-----S-D--	66
HPV16	-.....-CQ--NVC-D-I-TH--.N--T--R-H-DY--HM--CAIY----FKHIN----T--	66
HPV35h	-.....-M---Q--SVC-D-I--H--.T--TC-S-H-QY----I--CAVF-----KT-N-----TQ-I--	67
HPV31	-.....-Q--NVC-D-I--H--.N--R-C-H-DY--HI--C--M-----HSIN-----A-S--	66
HPV52	-.....-SIPA--N--V--I-D--.A--N--NA-----T-M-C--F--K-L--H-----M--	66
HPV33	-.....-EI-A--N--V--I-D--.A-KT--PS-----I-M-CA--T-KQ--FSH-C-----S-LA--	66
HPV58	-.....-EI-A--S--V--D--I-DI--.A-KN--TS-----I-M-CAIM--T--Q--SH-C-----S-VA--	66
RhPV1	-.....-M-A-A--S--DRI--.A--K-----CV-Q--CAV-----V-FSH-N-----S-T--R	67
HPV6b	-.....-AIAK--C--Q----.EN-T--HKHVL--CM-H-S--KQ--LSHI-M-----K--E	66
HPV11	-.....-AIAK--C--DQ----.EN-I--IHKH-M--CI--S--H--KQ--LSHI-L-----T--E	66
HPV44	-.....-IAKH--VC--Q----.EN-NK-TKH-Q--CI-Y-C--H--KQ--LNHI-M-----A--Q	66
HPV55	-.....-IAKH--VC--Q----.EN-NN-TKH-Q--CI-Y-C--H--KQ--LNHI-M-----A-T--Q	66
HPV13	-.....-IAKH--C--Q----.EN-NE-KKH-Q--C--Y-S--H--Q--LSHI-L-----T--Q	66
PCPV1	-.....-AKH--C--Q----.EN-NE-TKH-Q--CV-H-----Q--LSHI-L-----K--Q	66
HPV34	-.....-M--CK--S-C-DAI--.R--IH-S-H-D--HV-----H--LQSVNQ-A--S--R	67
HPV19	-.....-N---FNV--DQ-MNI--.SAAET--S--QI--K-A--F--RK-V--I-Y-P--T--E	66
HPV25	-.....-N---FNV--DQ-MNI--.TAAQT--A--QI--R-A--F--QK-V-----Y-P--A-M--E	66
HPV20	-.....-N--K-FN--DQ-MNI--.SAPDT--S-----QT--K-A--F--QH--S-V-Y-P--V--E	66
HPV21	-.....-N--D-FNV--DQ-MNI--.SAANTI--S--QT--K-A--F--QK--V--Y-Y-----E	66
HPV14d	-.....-N--D-FN--DQ-MNI--.TAANT--S-----QT--K-A--F--QN--V-----Y--T--I--E	66
HPV5	-.....-N---FN--DQ-MNI--.AAEQT--QA--K--QT--K-P-----Y--K-V-----Y-P--VK--E	66
HPV36	-.....-N---FN--DL-MNI--.AAEQT--A--K--QT--Q-A--F--QR-V-----Y-P--VK--E	66
HPV47	-.....-N---FN--Q-MNI--.AAEQT--KA--L--QT--K-A-T--F--QK--N--Y-P--A--I--E	66
HPV12	-.....-N---FNV--DQ-MNI--.AAEHT--T--A--T--R--A--Y--QK-----Y-P--T--E	66
HPV8	-.....-N---FNV--DQ-MNI--.AAEQT--A--A--L--K-A--F--QK-----I-Y-P-----E	66
HPV24	-.....-N-KK-F--V--DL-MNI--.QG-DT--S--QA--R-A--Y--QN-VL--YLP-----T--E	66
HPV15	-.....-D--F--N--N-MDI--.SGRD--I-T--L--QY--Q--O--F--F--KH--VM--DX-P-----T--E	66
HPV17	-.....DN---FNV--N-MDI--.SGQE--I-T--K--Q--Q--Q--F--Y--KN--VM--V--Y-P-----T--E	66
HPV37	-.....D--D-FN--N-MDI--.SGRD--T--M--Q--Q--Q--F--PHY--KN--VM--Y-P-----T--E	66
HPV9	-.....-A-FN--T-MD--.SGRE--QS--D--QT--Q--QI--HY--KN--VM--Y-P-----T--E	66
HPV22	-.....-K--FS--D--MD--.SGVE--T--Q--Q--Q--F--Y--RH--L--Y-P--T--T--E	66
HPV23	-.....-A--FS--D--MD--.SGLE--T--Q--Q--QI--Y--KR--M--Y-P-----T--E	66
HPV38	-.....-A-FTV--MDI--.SGVE--DT--Q--Q--Q--QIIYHY--RH--V-----Y-P--S--S--E	66
HPV49	-.....-A-NA-FNV--M-MDI--.SGKE--T--Q--Q--Q--A--FF--KHS--M--Y-P-----E	66
HPV4	-.....-S-VA-F----AI--THI-.SQEST--S--QY--ENI--K--AIMHY--KQ--L--K--L-PL-T--TE	66
HPV65	-.....-S-VA-F----AI--THI-.SQDDT--S--RY--ENI--K--AIMHF--KQ--L--K--L-PL-T--TE	66
HPV48	-QP.ETQ--S--T-FA-Q--IQ-T-I--.E-Y--K-HLAY--AV--I--A-Y--KEH--K--L-PL-T--TE	71
HPV50	-TQMETO--A-FL-Q--DIQ-N-I--.---N-K-H-DY--ESM-K-Q--AFY--KKENMS--L-PL--AK--E	72
HPV60	-N....QAD-T--S--QI--N--.Q--IQA--QY-D-N-KLY-TY-Y--KE-YSH--L-PL-A-Q--E	68
BPV1	-.....-AC--HVA--TQMQ--I--.S-DK-Q--H--LY-TAV--T--T--A--KK--V--CR--HSV-CQ	66
BPV2	-.....-AC--HVA--TQMQ--I--.A-DK-Q--H--LY-TAV--T--T--A--KK--V--CR--HSV-CQ	66
EEPV	-K....MSAA-DH-L-A--TQMQCI--.RL-Q--HACY-GAV-R-KL--A--TK-LKTI-CVP--CS-TA	68
DPV	-.....SAAK-Q--L-A--TQMT--I--.---T--K-H-DS-GPV-R-HG--A--HK-LIW--LNP--CS-KC	66
BPV4	-.....VS-EA-F--V--DQ--QV--.N--NT--LCLQY--A-I-R--A-Y-Y--QQ--K-----LYT--TR--E	66
HPV41	-S...QM-R-L--YI--QI--T--.V--H--RL--N--R--AIW-VL-QE-HA-V-GRA--AMT--E	69
COPV	-.....-K--A--L--E--S--.QN-QS-A--SR--S--K-Q--Y--GK--M--I-M-P--QS--Q	66
CRPV	-.....-A--Q--SI--E--S--.E--TS--S-LQ--N--K-Q--HFCKKH--RQ--YTP--S-LT-Q	66
ROPV	.....	0
HPV1a	-.....-N-S--L--Q-MN--.Q--LI--KQ--N--I-Q--Q--FHF--KN--VM--I-L-A--S--S--Q	66
HPV63	-.....-S-NN--W--Q--T--.I--MQ-N--Q--Q--FHY--KK--M--L--S--A--Q	66
MnPV	-.....-SIHS--A-V--E-MCM--.DGEET--A-LK--G--K-Q--HA--QH--HNKI--L-A--S-TQ	66

## E2 Appendix A

MOST-LIKELY AKAKQAIEMQLALESLQKSEYGTPEWTLQDTSLEMWLTPPKNCFKGGQTVEVIFDGDKDNAMEYTVWKYIYY	139
HPV54 G-GHK---L-----T---TV-S-----C--R-NA--TG-L-RR---D-----HQ--T-Q-VM-GD---	139
HPV32 ---HV---I-----T-LQ-TF-----E-Y---HAE-K-L-Q-R-----V---NPE---H-A-TF--V	139
HPV42 ---HM---IH---T-LQ-S---K-----E-N-L---N-K---Q-R-----KQ---H-A-T---I	139
HPV3 ---RS---VHVS-QQ---H-AHAQD---R---R---D-V---K-W---R-L---RY---ENK-C-VQ-RE-IV	139
HPV28 ---RS---VHV-LQ---E-A-AQDS---R---R---D-V---K-W---R-V---RY---ETKS-C-VH-RD-FT	139
HPV10 ---RN---VHV-QQ---E-A-AH---R---R---D-A---G-W---R-I---RY---ESK-C-VQ-REL-V	139
HPV29 ---CS---HV-QQ---Q-A-K---R---R---DAV-R-W---R-V---RY---ETK---CHVL---D-IV	139
HPV61 G---HK---VH-S-QG---T-A-AH---T-----N-Q-QR-W-K-RRLT-K---EDHK-V---VS-G---V	141
HPV2a ---C---V---QT-MQ-A-S-A---R-C---DA---K-W---K-S-L-K---SS-RD-I---S-GF--V	139
HPV27 ---C---V---QT-MQ-A-S-A---R-C---DA---K-W---K-LS-L-K---SS-RD-I---G-HH--V	139
HPV57 ---C---V---QT-MQ-A-S-A---R-C---EA---R-W---K-S-L-K---SC-RD-I---G-GH--V	139
HPV26 Q-W---IHI-Q---INTD-N-A-MR---Y-YM-E-H---E-T---T-V---CN-E-T-D-IR---V	139
HPV51 Q---C---HM-Q---N---D-NM---MRE-CY-L-CVA---Q---I-T---N---D-S-F---I	139
HPV30 ---CV---I-M---Y-T---KV-E---K-VCEN---H-A---Q---S-KRI---W---K---RT---V---QWV--	139
HPV53 ---CV---L-I---C-T-NM-E---R-VCES---Y-E---Q---Q-H---W---S---RA---V---WV--	139
HPV56 ---CS---V-I---STTI-NN-E---R-CE-L---E-K---E-HI---W---S-N-C-Q-VA---	140
HPV66 ---CS---L-I---AISNTI-KN-E---R-CD-L-R-E---E-HI---W---N-N-C---V---F---	139
HPV18 S---HK---L-M---QG-AQ-R-K---D---CE-L-N-E-TH---Q-Y---N---C-T-VA-DSV--	143
HPV45 S---HK---L-M---KG-AQ-K-NN-E---CE-L-N-E-SQ---K-H-Y---N---C-N-V---DS--	145
HPV39 C---Y---L-M---VAQT---N-E---K---N-L-H-Q---Q-T---WY---C---N-VL-GA---	144
HPV70 ---Y---L-M---AQTDFNK-E---K---N---Q-K---Q---K-V---WY---N---S-H-V---GA---	144
HPV59 N---CE---L-M---AQTFKN-Q---M-E-CQ-L-Q-A---K---Q-I---R-CS-E-T-H---S-TF--	143
HPV7 ---HA---MC---TT-NL---Q-L---AE---K---R-CNEH---H-L-TAV-V	139
HPV40 ---HA---MC---NT---NV---Q-L---AE---K---R-CNET---H-L-TTV-V	139
HPV16 N---L---L-T---TIYN-Q-SN-K---V---VY---A-TG-I---H-Y---Q---IC-T-H---N-TH---I	139
HPV35h ---M---L-M---T-NTT---S-T---E---I-LYT-V-QG---H-V---Q---Q-T-H---N-TH---I	140
HPV31 ---L---L-MM---T-NNT---KN-D---M-Q---LY---A-TG-L---H-Y---Q---VH-T-H---N-F--L	139
HPV52 ---C---L---A-N-TQ-S-DG---Q---RAE-QKY---H-Y-IT-QY-N---N-T-D---N-E---L	139
HPV33 T---F-V---L-M---T-S---Q-S-SQ---Q---V---CE-PK---Q-E---T-QY-N---K-T-D---N-GE---I	139
HPV58 T---F-V---L-M---T-NA-P-K-DE---Q---V---SE-QK---K-I---T-QY-N---A-T-D---N-SE---I	139
RhPV1 ---HK---V---N---NN-E---A---H-E---G---T-VP-T-L---C---T---VL-GH--V	140
HPV6b ---GHN---MH---LRT---SM---E-Y---Q---R---R-K---K---CAN-T-D-V---TDV-V	139
HPV11 T---GHN---MH---A-TQ---V---Y---R---Q-N---K---CE---V---V---TH---L	139
HPV44 T---GH---MT---T-LN-D---E---R---Y---Q---K---CNA---VW---V-V	139
HPV55 T---GH---MT---T-LN-D---M---R---A-Y---Q---KY-CNA---I---VS---V	139
HPV13 ---GHE---MT---T-LE---F-M---R---R---Q---KY-CNT---R-D-VS-T---V	139
PCPV1 T---GHE---MT---TVL---E-F---H---Q---RY-CNAE-S-H-VL---V	139
HPV34 S---GHN---L---NE-S-N-E---Q-W---Q-V---D---Q---K---RY-C---T-Q-V---TFV--	140
HPV19 ---E---G-V-Q-Q---S-V---A-TYRSA-E-Y---PMP---Y-K-A---NL---M---FV--	139
HPV25 ---E---G-V-Q-Q---F-K---S-V---T-TYKS---E-H---PMP---Y-K-A---NA---M-R---	139
HPV20 ---G-V-R-Q---S---S-V-A-A-TFRS---E-H---PIS---Y-K---NA---M-RFV--	139
HPV21 SR---G-M-Q-Q---AK---S-V---E-H---PVS---Y-N---NA---M-R-V--	139
HPV14d ---G-V-Q-Q---QF-S---S-V---G-TFRSA-E-H---PVS---Y-N---NA---M-H--	139
HPV5 T---E---A-V-Q---T-DFAH---V---I-TFRSA-EGH---PLP---Y-N-P---NL---M-T-V--	139
HPV36 ---E---A-V-Q-Q---T---AS-T---V---I-TFRSA-DGH---PVP---Y-N-A---NL---M-T-V--	139
HPV47 ---R-E---Y-V-Q---AFAL---V---T-TFKSA-E-H---PVP---Y-K-EA---NL---M-TFV--	139
HPV12 ----E---GIM-Q-Q---AS-N---V---A-TYNNV-EHH---PVP---Y-KEPE---NV---M---V--	139
HPV8 ----GIM-Q-Q---FAD---V---I-TYKNA-E-H---ATP---Y-KQP---NV---M---H--	139
HPV24 ----S-V-Q-Q---Q-P---K---V---I-TFKNT-E-H---PIN---Y-P---NL---M---V--	139
HPV15 T---D---G-VIL-Q---A-K---TQ---TVRSA-A---P-NI---M---K-P-E-I-V---T	139
HPV17 ----D---G-V-L-Q---P-K---TQ---TVRS---A---P-NI---M---N-PE-L-S---SF--	139
HPV37 ----D---G-VIL---Q-A-K-S---TQ---TVRS---A---P-NI---M---N-PE-L-V---A-SF--	139
HPV9 Q---D---G-V-L-Q---R-A-Q---AQ---AVRS---AYA---P-NI---VY---P-V-S---I-NF--	139
HPV22 S---D---A-G-L---Q-AE---VE---TVKS---AD---PKS---Y---PE-V-S---S---	139
HPV23 I---D---AIGIL---K-AD---VE---TIRS---VD---PK---Y---PE-V-P---S---	139
HPV38 ----D---S-V-L---K---AD-Q---AQ---AVRS---AD---PKNI---V---PE-L-S---T	139
HPV49 T---G-M-T-Q---PF-K-K---VN---TYNA-AQ---PYNI---PE-L-V---A-E-F	139
HPV4 YN---QIH-T-Q---L---PFAS-R---T-V-A-LIN-S-Q---L---YD-A-W---N-RQ---L---N-DFL--	139
HPV65 YN---QIH-T-Q---L---P-S-R---PEV-A-LIN-A-Q---L---YD-S-W---N-RY---V-N-D-L--	139
HPV48 Y---E---NI---LIQ---L-FAL-R---AE---A-TINSS-R---VPFI-N-W---N-ER-SFP---C-DF--	144
HPV50 Q---D---RI---L-Q---Y-DF-S---SEC---LNA---R---QPF---T-Q---N-PK-VYP-ICYE---	145
HPV60 Y---G-L-T---S---Q-AS-L-G-T---A-LL---R-T---K-Y---N-W---NNEN-TFP---N-E---	141
BPV1 ER---S-QE-S-T-F-D---S-L---WDRYME---R---ARV---E---NAS-TNW---YSNL-M	139
BPV2 ER---S-QE-S-T-F-N---C-L---WDRYME---R---ARV---E---NAS-TNW---YSKL-M	139
EEPV EQ---C---IV-E-LH-PWAK---S-T-L-W-RYQAA---G-L---ARV---EY---NSS-KTW---A-STV-V	141
DPV LE-R---LGN---KE-PWCN---S-C-L-WGRYQA---AETL---ARL---EY---SST-KTW---A-NSL-L	139
BPV4 Q---D---K-Y-C---FANQR-S-V---I-TFKA---E-TL-R---H-T---Y-QNAM-S-V---L---EV--	139
HPV41 -N-F---IK---KA-P-AA-G-S---E-TK-RY-AE-SRT---L---P-TLM---N-PE-LT-VVL---WV--	142
COPV -----QS-YID---LH-K-AN---C---R-RLVAE-AYT---KQID-RYGDSEE-IVR-VL-LD--	139
CRPV EC---V-YI---LR-P-SD---R-RFES---QKT---NPAT---YY---RG-NN---L-GIFI	139
ROPV EC-----V-----0	0
HPV1a E---T---V-H---KD-P---D-S---R-LF-A---AGT---S-S-L---TY-NNP---QTRH-I-NHV--	139
HPV63 D---T---T-Y-SG-RD-Q---S-Q-S---R-IF-A---DHT---I---Y-E-PN-STRH---RH--	139
MnPV QN---N---H-L-Q---AETP-AR-A---SQ---R-YMAG-SGT---D-TI-----T-M-T---K-GK---F	139

MOST-LIKELY	QD..D.DTWHKVEGGV р. TGLYYV..HGGFKTYVLFKDDAKKYG.....	KTGQWEVHGGEVI.....	193
HPV54	-NC.-GEG-T--CSNI-A.M-I-M..DAEH-V---D-KE-S-----	EY-----RM-SSI-----	195
HPV32	-TL.-G-C--Y-H-C..A---I..VDNM-QF-CN--NE-----	V-----D-TQ-----	194
HPV42	-TV.Q.G-C--Q-H-CH.A---I..VENM-QF-CN--EE-----	V-D-----D-NQ-----	194
HPV3	-NYT-.N-V--A-L-SH.E---M..-E-Q-F-K---RV-----	D--T-D----K-----HH.	197
HPV28	-NYS-.K-V--A-H-S..E---I..-E-EQ-F-K---YV-----	E--K-----K-----HH.	197
HPV10	-NYS-.R-V--P-K-S..E---T..-ENMNI--N---CV-----	E--K-----K-----HH.	197
HPV29	-NLS-.Q-V--K-Q-S..E---EDV-VF--K-HK--RV-----	E--I-----K-----HH.	197
HPV61	-ST.ET-L-Y--P-K-S..K---E..ME-QEH---T-AQE-Q-----	E--K-----M-NT--Y..EPC.	201
HPV2a	--TIT.-S---P-Q--E.L-----D-VRVN--D-GTESLT-----	V--T-----A-T-----HH.	197
HPV27	--AN----P-K--E.L---E..-D-VRVN--D-GTESLT-----	V--T-----A-R-----HH.	197
HPV57	--IN----P-Q--E.L--F..-D-VRVN--D-GIE-LT-----	V--T-----Q--R-----YH.	197
HPV26	KT...IG-C-GT-D..A-K-I-T..Q-AY-Q--D-QE-E-----	TGV-A--C-Q-----	193
HPV51	Y--N..K-V-TN-N---I-T..VNSK-E--Q--E--I-----	A.Q---YMY-T-----	192
HPV30	CG..-NG-T--PSV--.K-I---D-N-V--TD-N-E-V-----	YK-T-----M-N-S-----	193
HPV53	CG..E..G-C--SSA-S..E-I--I..-D-H--TN--E-T-----	CK-T-----M-KQS-----	193
HPV56	NG..CG-Q--CS--.R-I---D-H--TD-EQE--F-----	CKNI-----MEN-S-----	194
HPV66	NG..E.CG-C--SS--.R-I--M..-D-H--TD-EQE-----	C-NI-----MET-S-----	193
HPV18	MT..-AG--D-TATC-SH.R-----KE-YN-F-IE--SECE-----	N--T-----F-NN-----DCN	201
HPV45	ITE.T.GI-D-TAAC-S-W..V--I..KD-DT--Q--SECE-----	NSNT--QY--N-----DCN	203
HPV39	KN..NJ-I-C-T--C--W..I--M..NEHL-V--EV-IQ--ER-----	TS-K-----YN-NI-----HCP	202
HPV70	KTH.T..--C-T--Y--W..I---EQH---EV--Q--QM-----	TS-K-----CN-NI-----HCP	202
HPV59	VN..VGQ-C-TT-N--FW..--K..VEEEQV--K-IH-----	T-DK-----YN-K-----DCY	201
HPV7	-V..E..--T--Q--H.R--F-T..VH-CT----D-GKE-HT-----	--ND-T-I--SR-----	193
HPV40	-V..--A-T--K-Q--.K--S-T..VH-CT----D-AKE-QT-----	--NR-T-I--SH-----	193
HPV16	CE..E.ASVTT--Q--Y--E--IR-F-Q--E--S-----	--NKV--A-Q-----	193
HPV35h	LE..-SICTV-K-L-N..K-I---Q--VE--T-REE-----	--KNI-----Q-----	194
HPV31	CI..-GQCTV--Q-NC.K-I---E-HI--F-N-TEE-----	TGKK-----A-Q-----	193
HPV52	LG..E.CECTI--Q--Y--W..CD-E-I-F-K-SN--Q-C-----	V--V-----Q-----	193
HPV33	IE..E..--CTM-T-K--.I-M--I..-NCE-V-FKY--E--A-S-----	--QM-----Q-----	193
HPV58	IE..E..T-CTL-A-E--V--I..-NE--FKY--E--S-----	--QL-----SR-----	193
RhPV1	WG..-NG-V-TF-EA-NW..-H-T..VA-E-V--Q--YE-----	HGNNGNDGYE-----T-M..H.	201
HPV6b	--N..--V--HSM--A.K-I--T..C-Q--N--VKE-E-----	S-KH--CY-ST-----	193
HPV11	--N..-S-V--TSS--A.K-I--T..C-Q---N--NKE-Q-----	S-NH--CY-ST-----	193
HPV44	F..-T..-K-V--T-HI--K-----H---TN-EKE-E-----	NSL--CI-SSI-----	193
HPV55	H..-T..-K-V--T-HI--K-----H---TN-EKE-----	NSL--CI-SS-----	193
HPV13	F--T..-K-T--K-M--.K---I..-NL--LE-EKE-----	E-L--CI-ST-----	193
PCPV1	CE..N..-R-Q--K-M--I.K---M..V-QC--ID-EKE--Q-S-----	--L--CYDSK-----	193
HPV34	WL..E.GK-Y--SSH--.N-I--ET..QDNE-V--TQ-DR--R-----	VK-I-D-CM--K-----	195
HPV19	V-E..-N--S-S-NH..--FM..Q-N-RH--A--R-S-----	A--H--K-NK-TVF..TPVT	199
HPV25	V-D..-K--SAS--NH..-I-FM..--S-RH--A--RR-S-----	N--H--K-NKDTVF..TPVT	199
HPV20	--D..-K--SAS--NQ..-I-FM..Q-T-RH--A--SR-S-----	T-----K-NK-TVF..APVT	199
HPV21	V-D..--Q--SPS--NH..-I-FM..Q-T-RH--A--SR-S-----	R--H--N--N--K-NK-TVF..APVT	199
HPV14d	--D..-EQ--SAS--NH..-I--M..Q-T-RN--A--TR-S-----	--H--K-NK-TVF..TPVT	199
HPV5	M-A..-K--ARS--NH.I-I--L..Q-T--N--A--R-----	T--E--K-NK-TVF..APVT	199
HPV36	ME..-V--ARS--NE..-I--L..Q-T--Y--A--R-S-----	Q--K--N--K-NK-TVF..APVT	198
HPV47	M-S..-V--TTS--NQ..-I--L..Y-T--H--A--R-S-----	A--E--K-NK-TVF..TPVT	199
HPV12	M-P..E..-V--TTS--NQ..-I--L..-D--H--A-G-RM-S-----	--K--N--K-NK-TVF..APVT	199
HPV8	T-A..-K--TTS--NQ..-I--M..Q-S-RH--V-A--RR-S-----	A--E--KINKDTV..APVT	199
HPV24	M-D..N..-Q-Q-T-S-ANH..-I--L..I-E--H--A--NR-S-----	S--R--RINK-TVF..APVT	199
HPV15	-TL..-N--N--KI--.H-A--L..E-TL-V--IQ--EV--ARF-----	--I-----NEDT-F..APVT	199
HPV17	-NL..-N--R--.H-A--M..E-SL-V--IQ--EV--ARF-----	--R-----NEDT-F..APVT	199
HPV37	-TV..-N--H--.Y-A--F..E-DL-V--IQ--EG--ARFS-----	--R-----NKDT-F..APVT	199
HPV9	-TV.N..-E-Q-H--F-A--F..E-TV--IN-DK--AR-----	R--V-----NKDIVF..APVT	199
HPV22	-TD..-ES-E--H--.A--I..E-T---IK-ET--R-----	T--H-----NKDTV..TPVT	199
HPV23	-TD..E..-E--H--.A--FY..E-QL-N..IK-EA--RF-----	T--M-----NKDTV..TPVT	199
HPV38	LTD..E..-I-E--H--.A--Y..E-KL-V--KL-EN--R-----	V--L-----NKDTV..TPVT	199
HPV49	V-S..-M-Q--Q--E..A-A--K..D-TI-Q--T-A--VR-----	TS--Y--RINN-TV..APVT	199
HPV4	--M..N..EQ---K-E---D--FTD..-T-ERA-FT--SS--QRFS-----	R--L-T--FKTQ-----	195
HPV65	--V..N..EI---K-E---D--FTD..-T-ERA-FT--ST--HRFS-----	R--L-T--FKTQ-----	195
HPV48	--QNK--T--L--H..N-C--D..LN-DVF-FT--QP--V-----	--L-T-RFKNKT-----	200
HPV50	--R-K--K--L--H..N--FKE..VT-DSV-FK--QP--TV-----	--S--T-IFKNKT-----	201
HPV60	--IEQ--RTR-E--.N--FTE..NN-NRA-FL--DS--QT-S-----	Q--T-T--YKNQI-----	197
BPV1	RT..E..-G-QLAKA-A-G..-CTMA-AGRI--SR-G-E-ARFS-----	T--HYS-RDQDR-Y..AGV.	198
BPV2	RT..E..-G-QLAKA-A-G..-CTMA-AGRI--SR-GEE-ARFS-----	T--HYS-RDQDR-Y..AGV.	198
EEPV	RGTEE..EG-ETAVCAA-G.Q-I--CA..GMSS-V-FET-ET--RRWS-----	R--H-T-RDN.D--..YHST	201
DPV	RKP..-EEG-ETAT--A-A.D--F-TT..MS-TRV--E--ER--AR-S-----	T--T-T--RDNDRTY..HS.	198
BPV4	V-E..T..E---TSSDL--D-IF-ID..NQ-N-I--N-Q--AL-S-----	NS-MGQ--FESK-L-----	195
HPV41	ITP..T..-E-Y-AR--I-D..-I--ID..-ESV-M--R-DME-ENFS-----	E--TVTYRL-SALV-----	198
COPV	--EF--E-AH-KL-H..K--S-M..-TQQV--D-EEE-N--S-----	E--KY-ILNQPTT--.PTT	197
CRPV	GNA..-GE-V-T-S--.R-I--D..SE-NYV--D-ST--GRFA-----	AN-HYD-VFQNMRL-----	195
ROPV	.....#--PST--E.K-V--RD..TE-NNI--D-ET--ARFS-----	SK-EY--VYKSQKL-----	47
HPV1a	-NG..-V-R--SS--A.V-V--LE..-D-Y-N--AEE-S--S-----	T--YA-NYR-KRF..TNVM	199
HPV63	-NG..-NR-R-AASD--V.H-VF-LE..YD-V-N--D-QEE-NR-S-----	--RYT-QYE-KRF..TNV.	198
MnPV	A-P..N..GN-SRTTSHT-I..N-I-FN..KS-D-E--R--EE--R-S-----	L--T----D-L-THSLLIPVT	201

## E2 Appendix A

MOST-LIKELY	SSPPSVS.....	STGQAEV.....	STSE..TTSTLS.....	217
HPV54	F--A---.	--EE--LSI.....	-STG..-AEHTRP.....	221
HPV32	.V--A-I-.	--TTT-AEV.....	-S-G..L-ELVQT.....	221
HPV42	.V--API-.	--T..S.....	TDA..IP--G-T.....	216
HPV3	.D-FDP--.	--R..-I.....	PAPG..PLYA.C.....	218
HPV28	HAFDP--.	--R..-I.....	PAAG..PLC-GD.....	219
HPV10	.DAFDP--.	--R..-I.....	--PG..PVC--.	218
HPV29	.NAFDP--.	--Q.....	PATG..PLYASH.....	219
HPV61	A-VS-TQ.	DAV.R--.	--A..-AGH-RD.....	225
HPV2a	.T..A--.	--Q..-SA.....	--DD..PL-PIRT.....	220
HPV27	.T..A--.	--Q..-TA.....	--DD..PL-PIRL.....	220
HPV57	.T..A--.	--Q..-AT--.	--DDD..-L-P-RS.....	220
HPV26	CC-EF--.	--CSSNQI.....	--AK..-AEPV--.	218
HPV51	.TC-EY--.	--CSDLAL.....	--TT..-VEQ--.	216
HPV30	.YC-D--.	--LRSN-.....	-PV--.VVEYNT.....	218
HPV53	.YC-D--.	--FRSN-.....	-SV--.VNEYYS.....	218
HPV56	.YC-D--.	--CRYN-.....	-PV--.VNEYNT.....	219
HPV66	.YC-D--.	--CRYN-.....	PPV--.VNEYNN.....	218
HPV18	D-MC-T--.	DDT--.	--ATQ..LVKQ-Q.....	222
HPV45	D-MC-T--.	DDT--.	--ATQ..IVRQ-Q.....	224
HPV39	D-MC-T--.	DGS--.	P-T..L-TE--.	223
HPV70	D-MY-T--.	DDT--.	P-T..L-AE-Q.....	223
HPV59	D-MC-T--.	DEQ--.	--AG..SSEQ--Y.....	223
HPV7	C.--TV.	EGL.PI--.	APVD..IRHPAA.....	215
HPV40	C.--STIE.	EH-LPI-E--.	TADARPP-TATDT.....	221
HPV16	LC-T-F--.	--SN--.	--SP..IIRO.HL.....	215
HPV35h	VC-E--F.	--ST--.	L--A..IATQ-H.....	216
HPV31	VF-E--F.	--SD--.	I--FAG..IVTK-PT.....	216
HPV52	VC-A--.	--NE--.	--T--.AVH-C.....	214
HPV33	VC-T-I--.	--NQ--.	I--T--.ADI.QT.....	214
HPV58	VC-T-IP.	--DQ--.	I--T--.ADP.K.....	213
RhPV1	Y.-D--.	--ATHCDKL--.	P-V..IV-G-Q.....	225
HPV6b	C--A--.	--T.Q--.	--IP..S-TYTP.....	216
HPV11	C--A--.	--V.R--.	--IA..P-TYTP.....	216
HPV44	C--A-I--.	--V.QD--.	--IAG..PA-HS--.	216
HPV55	C--A-I--.	--V.QD--.	--IAG..PA-HT--.	216
HPV13	C--A--.	--V.Q--.	--IAG..PA-Y..-T.....	216
PCPV1	C--A--.	--V.Q--.	--IAG..P--H..-T.....	216
HPV34	CF.AP-F.	--PC--.	--P--.IVRP-HT.....	217
HPV19	--T-PD--.	PG--RDPN--.	TS-KTP--T-D-ASRLSPTASREQSQ.....	240
HPV25	--T-PF--.	PG--DSN--.	TS-KTP--D-A..RLSPTGSGERSQ.....	238
HPV20	--T-PD--.	PG--DSN--.	AS-QTPA-T-D-TTRQSPRKQSQQ.....	238
HPV21	--T-PD--.	PG--DSN--.	TS-TTPA-T-D-TSRLSSTRKQSQQ.....	239
HPV14d	--T-PE--.	PG--DSN--.	TS-KTP--A-D-TSRLSPADSRKQSQQ.....	240
HPV5	--T-PG--.	PG--DTN--.	T-PATP--.TAVDSTSRLQLTTSKQP.....	240
HPV36	--T-PG--.	PG--DTN--.	AS-KTS--T-.ATVDSTTKQLTTSEQP.....	239
HPV47	--T-PG--.	PG--TDPD--.	TS-KTP--T-AATDTSPRRQSINKQSQQETKRRGYGRPPSS.....	256
HPV12	--T-PG--.	PGQRDPD--.	A--KTPA--SD-TTRSSDKQSQQA.....	237
HPV8	--T-PG--.	PP--DTD--.	AA-KTPA--SD-TTRSSDKQSQQA.....	237
HPV24	--T-PD--.	PP--DTD--.	AA-KTPA--SD-TSRQQRSPAKQPQ.....	239
HPV15	--T-PD--.	PG-SR-L--.	PG-T..AN-KA-SPTQQPQACSDETTKRK.....	241
HPV17	S--AAGE.	GE-ATSID--.	--AP-SPANRQ--STVSSSRKRTTP.....	237
HPV37	S--AAGE.	G-DASPIN--.	AA-RSSPARG--ATSVSTRTTQR.....	238
HPV9	--S-PTG--.	G-DG-ASVH--.	TV-GSPLARGF-TTSVSTRKRT.....	238
HPV22	--T-P-G--.	DG-ETSKHT--.	L-R.GSP--R-PATTVPPTGGSR.....	238
HPV23	--T-P-G--.	VAS--QN--.	--AP..PA--SDSPQRSSQV.....	229
HPV38	--T-P-G--.	DASNNA--.	P-AS--LS-PQRSPSTN.....	231
HPV49	--T-P-G--.	DST.DSA--.	--RAALPEP--SVSPPERPPSQT.....	233
HPV4	--T-P-G--.	--PSTGLRESSNASPVHD-V.D-T--.	P--TTA--T-F-TTTATATATGAPELSSKTGTR.....	257
HPV65	--.I--.	--YSSSF--.	D-E-QQLPGP--T.....	220
HPV48	--.SV--.	--NTPSF--.	DFE-QQLPGP--T.....	220
HPV50	--A.SVT--.	--SR..NTN--.	PS--.SRVG--T.....	223
HPV60	--H.SVT--.	--SRS-FG--.	PAD-QPGP--SY.....	226
BPV1	--A.-VT--.	--SK-SSDD--.	Y--K..AQOQPH.....	221
BPV2	--TS-DF--.	RDRPDG-W--.	VA--GPEGDPAG.....	225
EEPV	--TS-DF--.	RDRPDG--.	--A--GPEGDPAG.....	224
DPV	FGA-PH--.	RNRDCIEGFW-DAG--.	ERRG--.	228
BPV4	H-A--H-RETIE.	GLWN-G--R-R--.	GR.P-N-PDR.....	229
HPV41	--SVT--.	--LR--.	G..-G--GGQRG-Q-GDH.....	219
COPV	NV-EP-T--.	V-D..SS--.	--R..R-PKVL.....	220
CRPV	--A.AGT--.	GPELPGH--.	--A-G..SGAC--.	219
ROPV	--.SVT--.	--PQ.PL--.	--AP..D-VPEEA.....	218
HPV1a	--VSSVT--.	--PLRPTIAL.GN-LDNA-A-PAP--.	.	76
HPV63	--TS-PR--.	AA-APA-H--.	DYP..-L-ESDTAQQST.....	230
MnPV	M--VN--.	--PLRTSGS--.	P-DTNPA-QGQ-T.....	225
	--T-QTG--.	FPRGDP--R--.	LHGN...--TG-PIPLRNSSSNQILL.....	239

MOST-LIKELY .....	TTTTTATQA .. RS .. RGTSEP . TRSRPR .. SRSRSR ..	248
HPV54 .....	.ANS .. -PR-DNS .. TK .. A .. I- .CTPP- ..	241
HPV32 .....	.TDLYN--P-P-T .. IT .. -SNCD- .DGTDGILYKDPTPTP ..	257
HPV42 .....	KLVQQVC .. --NPLH-T .. T-IDNHADCT .DGTAY .. NVPIQT-P ..	255
HPV3 .....	--QAP-- .. AQ .. V-A-- .. GPEQK .. -Q-LETVYG ..	246
HPV28 .....	--KAS-E .. T- .. V-AT- .. GPQQK .. -Q-LETLN ..	246
HPV10 .....	N--PAS-- .. AQ .. V-A-- .. GPEQK .. -Q-LEAV ..	245
HPV29 .....	N--RSP-- .. AP .. L-PE- .. GQE-K .. -R-LEAVGPGPQ ..	251
HPV61 .....	N--Q-T-TP .. TCV .. GP-QTSTSQT ..	249
HPV2a .....	AVSPVPAPV-AS-E .STG .. A-RAA-P-QALC .. -AQAPT-P ..	257
HPV27 .....	AVSPVPAPASAAS .. AR .. T-AP- .. NLLC .. TAPAPP-P ..	254
HPV57 .....	AAAAV-A-ATAT .. TA .. V .. P- .. LQDS .. AQAPS-P ..	249
HPV26 .....	.NA .. -Q-EAY .. VP .. V-K .. -EAY ..	240
HPV51 .....	N-P- -NPPL .. TC .. V-AK-AQ-QQ-K ..	239
HPV30 .....	YN-YQ-P-TS .. TP .. V-AN-A .AS-AR ..	241
HPV53 .....	HK-P-TST .. P .. V--Y-A .SS-LR ..	238
HPV56 .....	HK--T-S .. T .. V-NQDA .AV-HR ..	240
HPV66 .....	HR--TAS .. TF .. V-AQDA .AV-HR ..	239
HPV18 .....	H-PSPPSST .. V .. V-AKT .YGQTS ..	244
HPV45 .....	HAS-STPKT .. A .. V--PK .. HIQT ..	246
HPV39 .....	N--A-HST .. TT .. PC-QK .. -IPP- ..	244
HPV70 .....	H-PAH-A .. TT .. PC-KK .. -K-A- ..	244
HPV59 .....	PSA .. -PPE-Y .. LG .. P .. QTWN-QTK ..	244
HPV7 .....	-DA-D-KV .. HD .. APYAL- .ASTTK .. VYND-HAP ..	245
HPV40 .....	PDA .. PA-AT-E .. TV .. G .. P .. AQAP ..	240
HPV16 .....	ANH .. PAA-H-K .. VA .. L-E- .. -QTTI ..	238
HPV35h .....	.AYN .. E-H-K .. C .. V--T .. -QKTN ..	239
HPV31 .....	ANN .. --SNSKT .. CA .. L--GVR-ATT ..	241
HPV52 .....	--E-SK-S- .. V .. V-AKD .. -HLQ- ..	234
HPV33 .....	DND.NRPPQ-AAK .. -R .. -PADTT .DTAQ- ..	239
HPV58 .....	--EAT .. N .. N .. -S .. -QGTK ..	227
RhPV1 .....	HINPSPPP .. AN .. PSAK- .. NVWSS ..	245
HPV6b .....	AQ-S-L .. V .. SS-K-D .AVQT- ..	235
HPV11 .....	AQ--PT .. V .. AC-T-D .GV-A- ..	236
HPV44 .....	SSTT--LAQ-SST .. LP .. I--A-D .CVDA- ..	242
HPV55 .....	SSTT--LAQ-SP .. LP .. TC--E .RVDP- ..	242
HPV13 .....	--S-Q-ST .. V .. CSA-E .CVQA- ..	238
PCPV1 .....	--LAQ-C .. V .. SIAT-D .SVQA- ..	238
HPV34 .....	SNS .. SNA-D .. AG .. V .. P .. -K-H ..	234
HPV19 .....	QTNTKGRRYERRPSSR-RRQ-Q-R .. QK .. -SR-K .. SK--S .. -L-SN ..	286
HPV25 .....	QTSTKGRRYERRPSS .. R-RRQQA .. -QR .. -SR-KSRS--SQ .. -I-- ..	286
HPV20 .....	TNTKGRRYGRGPSS-RR-Q-RQ .. -R .. -SR-KSKSK--S .. -H-- ..	285
HPV21 .....	TNTKGRRYGRGPSS-RR-Q-HQ .. -R .. -SR-K .. S--S .. -L-- ..	284
HPV14d .....	QANTKGRRYGRGPSS-RR-E-RQ .. -R .. -SR-K .. S--S .. -L-- ..	286
HPV5 .....	QQTETRGRRYGRGPSS .. KSRRSQ-- .. QR .. -SR-R .. H--S .. -K-QTHTT ..	289
HPV36 .....	QQTETKGRRYGRGPSS .. R-RRPQAK .. QR .. -SR-RH .RS-S .. -Q--HT ..	286
HPV47 .....	RTRRPQTHQRRSRSSRSRSSQTHS .. --T-Y .. -S--LNK--A-S .. -T--T ..	314
HPV12 .....	DPRKGYGRGPSS .. R-RRQE .. -QR .. -SR-R .. Y--QS .. N----Q ..	277
HPV8 .....	QTETKGRRYGRGPSS .. R-RPQKE .. -R .. -SR-R .. H-T-S .. -L--V-AVGS ..	286
HPV24 .....	RYGRRESSPTDSRCRRSS .. RQKKQGRR .. -R .. -TR-R .. CS-TQ .. T---T ..	289
HPV15 .....	RTEARRYNRKESSP-- .. RR-K .. -Q .. GQRQ .. DTA-R .. -T-G-QE ..	282
HPV17 .....	TSPPRRYRRKASS .. P-A-TRHK .. -QD.I-- .. R .. S-TS .. G-QAISRG ..	279
HPV37 .....	PPRRYRRKASS .. P--A-R-K .. -Q .. G-- .. DTATR .. -T--GKQATSR ..	281
HPV9 .....	TSSRRYQRKASS .. P-RKKR .. -Q .. G .. E-EG .. GGEETNY .. R-Q-- .. KGRTET ..	287
HPV22 .....	THRYGRKASSP-I-- .. IRR-K .. -RE .. -QRO-TP--R-K .. T-----T ..	272
HPV23 .....	RRYGRKASS .. P-A-RR-K .. -Q .. G .. K-TL--R-K .. T----- ..	269
HPV38 .....	ARRYGRKASS .. PS--SRR .. -K .. GQRETTG-QR-RK .. -	273
HPV49 .....	KGRYGRKDSSPTAASNRSRKEVSRRSR .. RT .. R .. -EA-T .. S--QKA ..	310
HPV4 .....	SYSEV-EQASPT .. -R .. -KPRKSDAT-TT .. -PETEGV- ..	254
HPV65 .....	PTY .. -EL-QASP .. CG .. -K-RE .. SQPTSTT .. -PET-GL- ..	254
HPV48 .....	SSS .. SESPR .. -R .. PSI .. NSNTES .. TS-T--L ..	252
HPV50 .....	DKSQQRS .. G .. GQPQLQDTEP .. T-T-TV- ..	255
HPV60 .....	FFAASSS .. P--DGG .. T .. Q .. -GVSS-TT .. -P-AV-L ..	254
BPV1 .....	KEA .. EPAQPVSSL .. LG .. SPACG-IRAGLGW .. V-DGP--H ..	260
BPV2 .....	KEA .. EPAQPVSSL .. LG .. SPACV-IRAGLGW .. V-DGP-PH ..	259
EEPV .....	RGSDDTTRA .. LPYPA-R-SPIC-P .. VR-G .. N--AV .. H-QAPY- ..	269
DPV .....	AVL .. H-PPGGNTVH .. GP .. VRAC .. N-G-SI .. N-PTPY- ..	262
BPV4 .....	ARGRS .. RPSPRSSRD .. AR .. GRQQRAQSS-S ..	255
HPV41 .....	RPQGSRRR .. -N .. EE-G .. VAPAPK ..	242
COPV .....	L-PRKGPS .. -R .. P-RRSS .. RFP-R ..	240
CRPV .....	PDSAVPAQKKT .. G .. P .. KT-- .. TLG .. -R---PG ..	247
ROPV .....	AVS .. AGPAHHP .. TP .. V .. K-LSGPV-Y .. G-R---PGVG ..	111
HPV1a .....	SIDYTEL .. PGQGETS-V .. -Q .. -QQKT .. V-R--Y .. G-R---P ..	268
HPV63 .....	QTA .. RKAЕ-KGS .. -H .. HPK-PAVRKR--Y .. G-R---P ..	259
MnPV .....	REGRGDYPDGARRET .. RRYQQGPTP .. TP .. -SL-P-IY-PP .. -YEE--R-R ..	286

## E2 Appendix A

MOST-LIKELY . . . PRKR . . . . .	252
HPV54 . . . - - - . . . . .	244
HPV32 . . . - - - . . . . .	261
HPV42 . . . - - - . . . . .	259
HPV3 . . . EQQQQ . . . . .	251
HPV28 . . . WEQQ . . . . .	250
HPV10 . . . DGOH . . . . .	249
HPV29 . . . QQQQQQH . . . . .	258
HPV61 . . . H-- . . . . .	253
HPV2a . . . -A-- . . . . .	261
HPV27 . . . -A-- . . . . .	258
HPV57 . . . -P-- . . . . .	253
HPV26 . . . -G-- . . . . .	244
HPV51 . . . - - - . . . . .	240
HPV30 . . . -G-- . . . . .	245
HPV53 . . . -G-- . . . . .	242
HPV56 . . . -G-- . . . . .	244
HPV66 . . . -G-- . . . . .	243
HPV18 . . . AAT . . . . .	248
HPV45 . . . AT-- . . . . .	250
HPV39 . . . S--- . . . . .	248
HPV70 . . . SC-C . . . . .	248
HPV59 . . . TG-- . . . . .	248
HPV7 . . . - - - - - . . . . .	249
HPV40 . . . - - - - - . . . . .	243
HPV16 . . . Q-- . . . . .	240
HPV35h . . . H-- . . . . .	242
HPV31 . . . ST-- . . . . .	245
HPV52 . . . -Q-- . . . . .	238
HPV33 . . . LT-L . . . . .	243
HPV58 . . . - - - . . . . .	228
RhPV1 . . . -A-- . . . . .	249
HPV6b . . . - - - - - . . . . .	239
HPV11 . . . - - - - - . . . . .	240
HPV44 . . . -C-- . . . . .	246
HPV55 . . . -C-- . . . . .	246
HPV13 . . . -C-- . . . . .	242
PCPV1 . . . -Y-- . . . . .	242
HPV34 . . . - - - . . . . .	235
HPV19 . . . R-S-SKSRRKASTTR . . . . .	301
HPV25 . . . S-S-SRSESQSSKRRSRSSR . . . . .	308
HPV20 . . . S-S-SESPRRRSRYRSRSGSR . . . . .	306
HPV21 . . . S-S-SRSYSRSRSQSSDQPQYRFRSGG . . . . .	311
HPV14d . . . S-SQSSEERRSRYRSR . . . . .	303
HPV5 . . . RSTT-S-STSLTKTRALTSRSRS . . . . .	312
HPV36 . . . -TT-SATTRRSRSPSLAKTGQVQ . . . . .	307
HPV47 . . . STTS . . . . .	318
HPV12 . . . SQT-ALGATSVRSRSSRSPSVTQIRNR . . . . .	303
HPV8 . . . TTTS-S-SSSLTKAVRPRRSRSRSRG . . . . .	311
HPV24 . . . R-S- . . . . .	293
HPV15 . . . ISRGGN . . . . .	288
HPV17 . . . GER- . . . . .	283
HPV37 . . . GGDR-R- . . . . .	288
HPV9 . . . ERGGER- . . . . .	294
HPV22 . . . EQRG . . . . .	276
HPV23 . . . EQRG . . . . .	273
HPV38 . . . TNRG . . . . .	277
HPV49 . . . STS- . . . . .	314
HPV4 . . . L-R- . . . . .	258
HPV65 . . . V-RG . . . . .	258
HPV48 . . . RER- . . . . .	256
HPV50 . . . L-RG . . . . .	259
HPV60 . . . R-R- . . . . .	258
BPV1 . . . YNF . . . . .	264
BPV2 . . . -YHF . . . . .	263
EEPV . . . APSS . . . . .	273
DPV . . . TPQS . . . . .	266
BPV4 . . . -T-G . . . . .	259
HPV41 . . . - - - . . . . .	243
COPV . . . SGG- . . . . .	244
CRPV . . . VQR- . . . . .	251
ROPV . . . FD-ARS . . . . .	117
HPV1a . . . RGCG . . . . .	272
HPV63 . . . RDFT . . . . .	263
MnPV . . . KL-R-QDGRVKYPASPYRTKPPGETSSDDEDEGRGGHEPRPQRLPGLRGERAPERRRPPVQECEEDV . . . . .	357

MOST-LIKELY .....	PRGRGRSPS .. TTSSPSSQRS ..	RG ..	273
HPV54 .....	A-V .. Y- .. -DQQ H-T- ..	D ..	259
HPV32 .....	Y-QSL Q-P .. TKHL-HY ..	GVT ..	279
HPV42 .....	Y-QC-Q- .. OQLQH-NPSI ..	PSIP ..	282
HPV3 .....	QQQQQQQQ .. H-QT-AP-TT ..	E-A ..	273
HPV28 .....	Q-Q .. Y-Q .. T--T-TT ..	E-A ..	266
HPV10 .....	QOO-OQ-KD .. T-KA ..	AE-A ..	267
HPV29 .....	QQQQQ QQT .. P-HT--T-AC ..	A-T ..	279
HPV61 .....	Q-L .. HRD .. REQQ-DTTQK ..	DN ..	271
HPV2a .....	Q-VIV .. GQ .. QHPR-D-T-T ..	V- ..	280
HPV27 .....	Q-VIV .. GQ .. QHLR-D-T-T ..	V- ..	277
HPV57 .....	Q-VIV .. GQ .. QWQQ-D-T-K ..	VR ..	272
HPV26 .....	R-LS .. D .. -VTTVTTVTT ..	AATQP ..	267
HPV51 .....	Q-LTE .. D .. --TI- ..	PL ..	254
HPV30 .....	--TTE .. D .. --TDTT-Q ..	SA ..	261
HPV53 .....	--TTE PD .. -D-TTQ-TTT ..	AR ..	262
HPV56 .....	--L-E .. EF .. DS-RE-HAKC ..	VTTH ..	266
HPV66 .....	--ASE .. EP .. DS-RE-YAHC ..	V ..	262
HPV18 .....	-GHC- LAE .. K-HC ..	G ..	261
HPV45 .....	--QC- LTE .. -HH ..	G- ..	263
HPV39 .....	--QCA VTE .. PTEP ..	D ..	261
HPV70 .....	GVS .. R- .. E ..	TD ..	257
HPV59 .....	--QC- YTQ .. HP-ST ..	SV ..	263
HPV7 .....	R-D .. GD .. LSI-AVDGC ..	GRK ..	267
HPV40 .....	R-N .. GH .. LPITTTVGK ..	L- ..	260
HPV16 .....	--S .. E-D .. -GNPCHTTKL ..	LH ..	258
HPV35h .....	L- .. G .. -ELPYNPTKR ..	VR ..	258
HPV31 .....	--T .. E-E .. HRNTHHPNKL ..	LR ..	263
HPV52 .....	R-PDV .. TD .. SRNTKYPNNL ..	L- ..	258
HPV33 .....	FCA .. D .. P ..	A ..	249
HPV58 .....	R-LDL PD .. RDNTQY-TKYTD ..	CAVD ..	252
RhPV1 .....	V-RSDSGGD .. P ..	V- ..	261
HPV6b .....	A--VQQ-C .. NALC ..	VAH ..	255
HPV11 .....	A-- .. - .. TNNTLC ..	VAN ..	254
HPV44 .....	-- .. P-T .. N-NNARNTVC ..	V-N ..	265
HPV55 .....	-- .. PTT .. N-NNARDTV ..	V-H ..	265
HPV13 .....	Q--PS-PIG .. NPQNTQ-IVC ..	VTD ..	264
PCPV1 .....	L--PSHCAR .. KLQNT-NIVC ..	ATD ..	264
HPV34 .....	QCD .. -D .. E ..	G ..	242
HPV19 .....	GRGRGSPTATSDQSS .. A--TT-L- RGSSRVG ..	RSRGGRSRV ..	347
HPV25 .....	KTSATRGRGPSPTTTSD-AAR .. -- .. AT--QRSR ..	SRAGSSR ..	355
HPV20 .....	GRVALRAITTNTTTTR-AG-G-T .. S--TT--RQ ..	LRGGGR ..	349
HPV21 .....	QVSLITTATTNTATNYSTRGSG--SS .. -T .. SST-KRPR-P ..	-- ..	354
HPV14d .....	.SRKEVSRITTTT--GS .. S--KR--A ..	-- ..	336
HPV5 .....	RGRSPITCRRGGGRS--R-S-- .. S--CTT--QRARAESST .. TRGARGSRGS--	367	
HPV36 .....	RVSTRSRSRSTSRRGGR-R-S-- .. S--TTTTNKRSRVRAETT .. GSRGARGGRGA--	364	
HPV47 .....	R--GR GS .. -RQRSR-PSTYTSKRSREGNTRGRGRQGRA--	359	
HPV12 .....	RSRSQSRGRGGRGSSTDTTTKR-RSR .. S-- .. SNTRKR--G ..	ERGR ..	346
HPV8 .....	RATATSRRRAGRSPRRR-STS .. -NTFKR--GG ..	GR-- ..	352
HPV24 .....	STS--NRRC .. RGDT-RG--G ..	VS ..	314
HPV15 .....	Q-R--RE .. -SI--AWG-G ..	GRSR ..	311
HPV17 .....	Q-R-E--Y .. RD--R-PN-G ..	RG- ..	305
HPV37 .....	R-E-SY-RD .. -S--DRG-G ..	GR ..	309
HPV9 .....	R-- .. SS-AD .. S-TPTDRR-G ..	-- ..	315
HPV22 .....	G-ATR--L .. RE-AE-PR-G ..	GRG- ..	299
HPV23 .....	G-ETQ--S .. RSA-K-PR-G ..	GRS- ..	296
HPV38 .....	G-DTR--S .. RG--V-PT-G ..	--R ..	298
HPV49 .....	GSRGSGGSV .. --RD--PKR ..	TR-- ..	337
HPV4 .....	R-EGKSG-G .. SGET-RKR-R ..	GG- ..	280
HPV65 .....	R-Q-KSG-G .. PGET--KR-R ..	GGGR ..	281
HPV48 .....	R-EPRE GT .. -DTT-RRRG ..	KRKL ..	279
HPV50 .....	R-E-E HH- .. YRHRK-QSEL ..	GAD ..	280
HPV60 .....	SNEQQ-EL .. SRE--RTK-RRV ..	PDEVDR ..	285
BPV1 .....	-A-S-G-II .. RS--TPV-GTV ..	PVDL ..	288
BPV2 .....	-A-S-G-LL .. RSA-TPV-GPV ..	PVDL ..	287
EEPV .....	-GSSV.G-D .. SP-ES-R-VP ..	LVLL ..	295
DPV .....	--SGV.G-D .. --PLP-PVEQNPR ..	CVSLPD- ..	295
BPV4 .....	-HSS-..RD .. -RLPSPGRPP ..	GG-R ..	280
HPV41 .....	R-- .. AYG .. RR--KA--R ..	TA ..	261
COPV .....	G-LGR GG .. GELP-QP-P-S ..	SWSP ..	267
CRPV .....	--AKQRKQAA .. PDEAD-AAGD ..	I-P ..	273
ROPV .....	R-QGKHPTDFNAN-I-AD-TD ..	TD ..	141
HPV1a .....	R-EGE .. TP ..	--T ..	283
HPV63 .....	L-RGE GE .. ARA-AG-GER ..	VAF ..	284
MnPV .....	DGVGALLDDLKLYQEPPGDPVEDSDSPGS-LTP.A-P .. DL-RYD-T-L ..	QVDA ..	407

## E2 Appendix A

MOST-LIKELY GGPVDSGRRRSRSSSTS.SSNKR.....	RLGRLLDEARDP	307
HPV54 .--GCNDN-HI-.DDN.NK-QG	.H..TSSGD.TT	287
HPV32 NV---P-SQ-.VT.-DN.NN-Q-	.N..PC.GNQTT	308
HPV42 SAS---P-LCGV-TN-EN..C---	.N..HC.GSQAT	312
HPV3 RQ-L-TD-T-D-D..T-C..PHPI	.GH..RS-PD.CV	302
HPV28 SQ-L-VT-TSDCD..T-C..PYTV	.GH..PS-PD.CA	295
HPV10 --Q---D-T-LCD..TR-.AHPV	.H..PS-PD.CA	296
HPV29 ----N-T-DCD..----Q-PY	.H..PS-PD.CA	308
HPV61 HKR---TDQW.INGHRN..TETG	.DN..C-SY.SS	300
HPV2a E-E-ECYNK---I-.DSN.RTDP-	.WG..HG-TD.SV	310
HPV27 E-Q-ECYDK---I..N-N..TAP-	.WD..HG-TD.TV	307
HPV57 E-Q-ECQND---IRNPD..TDP-	.G..HS-LD.AV	302
HPV26 -QS-YTNNNLH-T-GG..HHPG	.D..TS.SDQTV	297
HPV51 ..S-NTNNQIHCG-G..T-TG	.GH..QS.ATQTA	282
HPV30 ..ARE-HAN-VNTNN-N..NRQC	.LGGATCYNTEVDGGYKTT	298
HPV53 ESYAECVA-N.TD.N-N..N-T	.KHLPGGASCNNTEIDSGYKTA	302
HPV56 THIS-TDNTD---R-IN.NN-HP	.GDKTT	293
HPV66 ..TT-TDISN.NAN-R-PRI-TQ	.SH..C.GDKTT	290
HPV18 .--NELLGA.AT..P-G..N---	.K..-CSGN.TT	287
HPV45 ..VNTHVHNPLLC----.N--	.K..VCSGN.TT	291
HPV39 -VSL-HLNNPLH-N--G..H-T-	.Y..-SCGN.TT	291
HPV70 -VF----.LVT..KG..C---	.H..QCCGD.TT	281
HPV59 ..S-YCDNPVVRLHPG..N-P-	.H..IPCSN.TT	291
HPV7 ..Y-T-N-A..-PDIE..N--I	.N..SGGGHST	295
HPV40 -EY---TAD-T..-TPDPE..-N-GH	.N..CGGGST	290
HPV16 RDS---.SAPILT.AFN.--H-G	.I..NCNSN.TT	286
HPV35h LSA--VD-GVY-T-DC..T--D	.C..GSCST.TT	288
HPV31 -DS---VNCGVI..AAA.CT-QT	.A..VSCP..TT	293
HPV52 QQS---TT-GLVT.A-E..CT--G	.V..AHTTCTA	288
HPV33 .L..N-TA-T.A-N..CT--Q	.T..VCSSN.VA	274
HPV58 ..SRPR-GGL.H..T-N..CTY-G	.N..VC.SSKVS	279
RhPV1 ..AL-GKS-SVLCG-AH.NNATG	.SSGD.....SD.YT	289
HPV6b I---HNHLITNNHD..QHQ-	.N..NSNSS.AT	285
HPV11 IRS---TINNIVTNDYN..KHQ-	.N..NCHS..AT	284
HPV44 SDS---TNNNILPN-YN..---G	.D..NNYCT.AT	295
HPV55 SDS---TNNNIYPN-YN..---G	.D..NNFCT.AT	295
HPV13 YDTL--ANNNINVNHYN..N--G	.D..NSYC..AT	294
PCPV1 R-TL--ENNI..NNNNYN.NN-QQ	.N..NSNSS.GT	294
HPV34 .-L-.FVHNLQPT.TD..-TQC	.T.....HN.VA	266
HPV19 RSRGRGK-S..E-P-PT..NT--SRRQSGSS.RLHGVSAD	.AVGTSVHTVSGRNTG-----E--L--	410
HPV25 --RGRG---H-L..E..PTS--SRRESGSV.RLHGVSAD	.AVGTSVHTVSSRHTG-----E--L--	419
HPV20 --SRQRA-G-----PTPS--SRGESESV.RQHGSPS	.DVGTAVYTUVSSRHTG-----L--	414
HPV21 -AIGG-SG-GR-----P-PS--SRGKSESV.RQRGISPD	.DVGKSLQSVSTRNTG-----L--	420
HPV14d R-RGG..S-G-----PT-S--SRRESESS.RQRGISPS	.DVGKSLQSVSSRNTG-----L--	400
HPV5 -SRGGR---G---S-SPAHS--SRGSAKL.RGVSPGE	.VGGLSLRSVSSKHTG-----E--	431
HPV36 -SGGRR-G-S..PAH--SREHSVRS.RGVSPDQ	.VGKSLRSVSSKHTG-----E--L--	426
HPV47 SSGGREQ---R-F--PD-S--VRRESPKY.RGVSPSE	.VGKQLRSVGAHKSG-----E--	423
HPV12 -ERGGR-K-----PTP--SRAGSRSS.RQRGVSPSE	.QVGRSLQSVSSKHTG-----E--L--	411
HPV8 R-RGSR-----E-----PTPT--SRGESSRL.RGVSPSE	.QVGRSVQSVSAKHTG-----I--	415
HPV24 TSSRGR--GSR-----S..-PTP-TKAQRGC.DTRSVRDGSIS	.PGDVGRKLQTVSGRQNSG-----E--L--	384
HPV15 R--TTRSQSKE-L-R-R..R-KS-SRGSSPR..GGISPAD	.VGSSVRSLSLGRKHTG-----E--E--	373
HPV17 SSGGPTT-SQ---L-R..R-RS-SRSRGSSAGGGVAPEQ	.VGKSVRSVGRNPEGG-----T--E--	369
HPV37 SRGGPET-SQ---L-R..R-RS-SRGSS..SR.GGVAPDA	.VGKSVRTVGRDHSG-----K--	371
HPV9 --RGPTT-SQ---R-R..H-RS-SRGGTASR.VGVSPDE	.VGTRVRSVGAHHG-----A--A--K--	378
HPV22 ---LTRS-S---RTRB..-VDGG.....GVAPDE	.VGATLRSIGRQHSG-----AQ--A-K--	353
HPV23 ---LTRS-S----.PE..VTGG.....GVAPSE	.VGASLRSLRSRHSSG-----AQ--A-K--	348
HPV38 R-GG--R--G.PV..TR..R-RSLSRASSAG..GGISPDK	.VGTAVRSVGRQSGG-----T--AD-A--	358
HPV49 R-RGGRS--S.PT P--..TS--ERRRSRSR.GGEPVSGVGIVISPDKVGSRVQTVSGRHLG-----E--S--	405	
HPV4 R-GGETELGSAP-PAEV.G-RH-----	.QVERQGLS---L-QA-----	322
HPV65 --GETLESA.P-PGEV.GIRH-----	.TVERQGLS---Q-QA-----	322
HPV48 -SDSAPTPEVG-R--T.LARHG	.YS-----QE-----	315
HPV50 SA-TPEEVG..--.H-V..AAHG	.LS--R--QE-----	313
HPV60 QSA-G-APTAEEVG--RH.R-LP-SGIS	.A--A--QG-----	323
BPV1 ASRQEEEQ.S.PD.--E..EEPV	.TLPRRTTNDGFH--.K-GGS	326
BPV2 APRQEEEENQ-PD.--E..EEPV	.TVPRHTSDAD-FH-LK-GQS	327
EEPV P--S-PAPPS.PD-TDV.IAEGDKEPE	.FSI-SKPG.GQ	331
DPV F-RGEEDNPP--PDQHDV.IP-PQPKEP	.F..S-FGSSGGL	332
BPV4 R-TPERE-CP.GT..P-P..PTPD	.Q.....VGGRSSTPKRQASS--AQ-I-A-Y--	326
HPV41 AS--SR-NGG-SD.F--..GESD	.EGHRVRH-AL-KK.T-GVA	299
COPV PS-QQV-SKHQLR.T---.AGG	.Q--Y-----	299
CRPV PA-E-V---TTTVGR-P.PGRN-	.RE-IT--S--	307
ROPV FS-AFRPPT-EVGRRN.TTAP-	.ESARGLG-VRQ-IS-----	183
HPV1a P-S-P-A-DV.G-IH-T..PQ-G	.HSS--R--Q--W--	318
HPV63 IS-G..-VGT-TR..PP-G	.GQS--R--IQ-----	315
MnPV ESSPPRT-PAPTLVAE.CTPG-----PSPQT	.GSGQQALGEPPSRPS-G..HCRDP-TA	459

MOST-LIKELY P.....VIHLKGDA	NTLKCFRYRLKKKYKGLFKNISTTWHWVG	GD..GT..ERLGI..VTITFTSETQRQDF	368
HPV54	-.....IV-F--EP-----Q-IQ-.--H--EQA-S---ACVP...-T.KNR--..-L-YS-VE--Q-	348	
HPV32	-.....Q--P-C---L-W---NCSH--TQV-S---LTEK...Y-RDSKD--.I--HYYN-E--DK-	371	
HPV42	-.....Q--P-C---L-F--RNCSH--TQV-S---LTEN...C-RDTKT--.I--HYYD-A--NL-	375	
HPV3	-.....R--P-C-----N-GKNK-YSRT-S--R-S.CE..SE..NQCAY...WY--YG--EA-	362	
HPV28	-.....V----P-C-----H-GKRK-YCKT-S--R-S.CE..SE..NQAAF...WY--YS--NE	355	
HPV10	-.....R--P-S-----HHGKRK-YSRS-S--R-S.CE..SE..NQAAF...LWY--D--TE-	356	
HPV29	-.....R--P-S-----QNGK--YCKA-S--R-SCEP..EN..QS.AF...WY--V--AE-	368	
HPV61	-.....P-K-----QHSVPE--DKA-S--A-Q...S..T-AAF...LWYVNVE--KQ-	361	
HPV2a	-.....R--C-----VQ-HKDV-YARV-S--A-N...D..KT.AF...LWY--VE--TE-	370	
HPV27	-.....R--C-----VQ-HKDK-YDRV-S--A-K..CD..KT.AF...VWY--VE--KE-	367	
HPV57	-.....Q-E--C-----VQ-HKDV--VKA-S--AC-N...-D..KT.AF...LWYK-QE--AE-	362	
HPV26	F.....IV----T-S--L--F--H--YC-V-S--TSN..TN..QQ...N-I--NN	356	
HPV51	F.....IV----T-C-----FT--H--Y--V-S--T...SN..TKT--..V-D-AH--ET-	339	
HPV30	-.....V--EP--R--L--CQ--H-H--V--S--Y--T.NT..H--Y.SY..I--VVYKD--AN-	356	
HPV53	-.....V-I--E--R--L--FQ--H-Q--VTV-S--Y--TNV..CA..VNNSY..I--VVYKD--K-	362	
HPV56	-.....V--EP--R--C--FQ--T--VDVTS--Y--TST--..NK..NY.S--I--IYKD--NS-	352	
HPV66	-.....E--R--C--FQ--T--TDVT--Y--TST--..NK..DS.S--I--LYKD--DT-	349	
HPV18	-.....I--R--S--L--R--HSDHYRD--S--T--A--..N..-KT--..L--V-YH--TK-	346	
HPV45	-.....I--K-S--L--R--ADHYSE--S--T.G..CN..KNT--..L--V-YN--V--NT-	349	
HPV39	-.....I--K-G--L--Q--..DT--E--C--IR-K--..KNA--..L--V-YAT--S--K-	351	
HPV70	-.....IV--K-G--L--R--FNS-YE--C--I--K--..S..KHT--..L--V-Y-T-A--K-	341	
HPV59	-.....I--K-G--L--R--VHW--E--S--T--NR--..S..AKT--..L--L--Y--NE-	351	
HPV7	-.....I-Q-E--C-----T..VSH--YT--S--R-TTES..R..NKN..I--L--YS-VH--SQ	355	
HPV40	-.....I-Q-E--E--C-----G..VSH--C--S--R-TTES..R--..KNA..I--L--YS-VQ--S-	350	
HPV16	-.....IV-----L--F--HCT--YTAV--S--T--HN..VK..HKSA--..L--YD--W--DQ-	346	
HPV35h	-.....IV-----L--G--..A-YQDA-S--R-TCTN..DK..KQIA--..L--Y-T-Y--DK-	348	
HPV31	-.....I-----I--L--S--..Q--YEQV-S--TCT--..K..HKNA--..L--YI--TS--D--	353	
HPV52	-.....I--P-S--L--V-T.H-S-YVQ--S--TSNE..C--N..NK--..YSD--Q-	349	
HPV33	-.....IV--ES--S--L--P--E--YSSM--S--TSDN..KN..SKN--..V--VT--Q--M-	334	
HPV58	-.....IV--P--S--L--P--F--D--YC--M--S--TSD--..KG..DKV--..V--Y--T--L-	339	
RhPV1	-.....IV--ES-C--L-F--G--H-H-YI--S--R-A.NH..AS..-K.A--..V--AN-L--Q-	347	
HPV6b	-.....IVQFQ-ES--C-----NDRHRH--DL--S--ASSK..AP..HKHA--..V--YD--E--Q-	346	
HPV11	-.....IVQ-Q--S-C-----ND--H--ELA-S--ASPE..AP..HKNA--..L--YS--E--Q-	345	
HPV44	-.....VQ-Q--C--L--HA--T--VAA-S--R-TCS--..TS..SN.AL--..L--YVD--Q--Q-	355	
HPV55	-.....VQ-Q--P-C--L--HA--H--T--VAA-S--R-TCS--..TS..SKHAL--..L--YVN--E--EQ-	356	
HPV13	-.....IVQ-Q--S-C-----HE--D--LLA-S--TAPN..NS..OKHAL--..L--YVN--Q-	355	
PCPV1	-.....IVQ-Q--S--N-----HD--H--MLA-S--TASS..NS..TKNA--..L--YVN--Q--	355	
HPV34	-.....IV--K-S--L--MH--G--SH--N--VT--..T..NN..TN..SKC-V..I--FM-S-TS--QKQ-	326	
HPV19	-.....LVR--EP--RS--N-A--HM--R--SSF--A-S--A--..I--..RTRML--S--V--FN--KH-	473	
HPV25	-.....LVR--EP--RS--N-A--HM--T--SSF--A-S--A--..I--..RSRML--S--I--NS--KH-	482	
HPV20	-.....LVR--EP--N-A--QR--T--Y--SF--A-S--A--..I--..RSRML--S--I--FS--K--	477	
HPV21	-.....LVR--EP--N-A--L--Y--AF--A-S--A--..I--..RSRML--S--F--FE--K--	483	
HPV14d	-.....LVR--P--R--N-A--Q--FT--YRAF--A-S--A--..I--..RSRML--S--F--FN--R--	463	
HPV5	-.....IV--A--NV--N-A--I--M--RSF--S--A--..I--..RPRML--S--S--Y--R--	494	
HPV36	-.....LVR--E--N-A--I--M--YRSF--S--A--..I--..RPRML--S--S--YN--R--	489	
HPV47	-.....LVR--E--N-ARN--R--RSF--FS--A--..SI--..RSRML--S--SCL--R--	486	
HPV12	-.....IC--G-----N--ARH--T--AF--S--A--..S--..RPRML--S--TN--K--	474	
HPV8	-.....LVR--E--N--ARVR--R--YF--S--A--..S--..RSRML--L--AG--K--	478	
HPV24	-.....L--R--G-----N--A--LR--R--HY--AF--S--S--AA--..RSRLLVS--FK--SG-	447	
HPV15	-.....L--R--K--F--A--..QD--V--YY--S--T--..SN..D--I--RSRLLA--S--N--E--EL-	436	
HPV17	-.....L--R--E--K--A--R--GS--V--YY--S--AN--..TN..D--I--RSRMLLA--NTYDE--EL-	432	
HPV37	-.....V--R--K--Y--A--HGN--V--YY--S--S--..TN..D--I--RSRMLLA--Q--N--E--EL-	434	
HPV9	-.....LML--R--V--Y--F--ER--KR--V--YY--S--S--..SC..D--V--RARMILA--DTYEH--Q-	441	
HPV22	-.....L--R--A-----Y--FR--HA--S--QF--S--H--..T--D--I--RSRIL--S--HTDRE--EKC	416	
HPV23	-.....L--R--G-----Y--FR--HA--K--YYV--S--I--H--..S--D--V--RARML--A--H--NHE--EKC	411	
HPV38	-.....L--R--Y-----Y--FR--HA--G--RFV--S--I--DA--..SN..D--I--RSRMLLA--Y--S--EK-	421	
HPV49	-.....L--R--P--I--Y--D--RKL--V--HY--S--V--..N..--I--RSRMLLS--NST--SQY	468	
HPV4	-.....M--L--T--S--W--KVNSNCCN--LFM--V--N--..CS..HNHSR..ML--A--D--TD--DA-	382	
HPV65	-.....M--L--T--S--W--KQNNSNCG--LFM--V--N--..VS..NHSR..ML--A--K--PG--DS-	382	
HPV48	-.....LVLFT--QQ--N--W--N--CTT--AS--LCF--SV--K--L--PN..SD..GGAAK..LVA--K--DA--V	376	
HPV50	-.....LIIT--QQ--N--W--FSQ--AD--YECC--SA--K--L--PK..SE..GYR--DAKLL--A--KNPE--LS-	376	
HPV60	-.....ILLI--L--S--W-----TRY--CM--VFR--DI--..VP..S..SRHKLVV--NDT--DV-	384	
BPV1	C.....FALIS--T--QV--Y--F--V--NHRHRYE--CT--FT--ADN..-A..-Q--QAQL--G--PS--	389	
BPV2	C.....FALIS--S--QV--Y--F--V--NHRHRYE--CT--SFT--ADN..-A..-Q--QAQL--G--PG--	390	
EEPV	-.....CLI--S--NG--QA--Y--F--C--RYFREHYQH--T--WT--ER..-S--H--DAC--LV--KDSS--GV-	394	
DPV	-.....CLLIS--TG--QV--YSF--V--RWHDRKYHHCT--WA--EQ--..S--P--DAT--IV--KDQS--SM-	395	
BPV4	-.....LL--Q--A-----R--ATQAPHK--LCM--S--T--SKT..SP..LKS--H--RML--A--SNSE--NC-	388	
HPV41	-.....AEGHLYVGA--PV--S--R--L--KW--N--S--DIMYLG--FT--TES--..C--SGRFFCA--SN--K--EK-	367	
COPV	-.....LV--A--P--S--I--SH--HR--YLGA--K--TS--GDGASK..HDR--SARMLLA--L--DQ--E--	365	
CRPV	-.....C--GH--Q--L-----S--HSS--DC--S--DTT..S--C--SGRML--K--ADSE--DK-	370	
ROPV	-.....C--GN--Q--L-----A--HRT--DC--S--DNS..S--C--V--SGR--L--K--KD--A--EKV	246	
HPV1a	-.....VCV--G--Q--L-----ASTQVD--DS--TDRK..N--I--SARMLVK--ID--A--EK-	381	
HPV63	-.....I--C--GP--Q--L--I--ASNSSD--ES--HNK..C--D--V--HARMLVR--I--TE--DR-	378	
MnPV	C.....LLII--SS--QV--L--F--SWHHS--SY--Q--PSV..S..N--I--RSRILVMCEDSA--MDR--	522	

## E2 Appendix A

MOST-LIKELY	LNTVKIPKGVQVSLGYMD...	SLG	389
HPV54	-V--R--PSISM--V....-\$		367
HPV32	-S---L-P-IKSCI---SMIQLFMS		394
HPV42	-----S-IKSCI---SMIQLFI\$		398
HPV3	-S---V-P-I--I--H-SM..FT\$		383
HPV28	-S---V-P-I--I--H-SM..FV\$		376
HPV10	--V--V-P-I--I---S...IFS		376
HPV29	-AN---P-M-AI--H-S...VF\$		388
HPV61	--R-T---I-ATA---SM..CI\$		382
HPV2a	-TR-S---LIALP---SA..FV\$		391
HPV27	-TR-N---IALP---SA..FV\$		388
HPV57	-TR-HL---KALP---SA..FV\$		383
HPV26	-T----QSITST--I-----\$		375
HPV51	IK-I-V-PS-TL---I....T-\$		358
HPV30	--V---PSIKIVM-H-TGV.DM\$		378
HPV53	-DI---PS-SLV--H-TCV.DM\$		384
HPV56	-SH---#S-----Q-#.....		368
HPV66	--V---PS---I--Q-S...CP\$		369
HPV18	----A--DS--ILV-----TM\$		365
HPV45	-DV-T--NS--I-V-----TI\$		368
HPV39	-D----SS-H-----T-\$		370
HPV70	-E--R--PS-H--V-----T-\$		360
HPV59	-D----NS--IHV-----V\$		370
HPV7	-AL----TIKH---MLT...IM\$		375
HPV40	-AI----TIKH---MLT...LM\$		370
HPV16	-SQ----TIT--T-F-----I\$		365
HPV35h	-T----NT-T-K-----I\$		367
HPV31	-----NT-S--T-----TI\$		372
HPV52	-K----NT--IQ-V-----\$		368
HPV33	-G----PT--I-T-F-----T-\$		353
HPV58	-----PT--I-T-V-----\$		358
RhPV1	-----ST-TL-Q-V-----TV\$		366
HPV6b	-DV---PTISHK--F-SLH.L-\$		368
HPV11	--S---PTIRHKV-F-SLH.L-\$		367
HPV44	----L-PK-TYKV---SLQ.L-\$		377
HPV55	----RL-PT-TYKV---SLQ.L-\$		378
HPV13	-K----PTITHK--F-SLQ.L-\$		377
PCPV1	-----ATIKHT--F-SFQ.L-\$		377
HPV34	-QCA---PTIS--S-----I\$		345
HPV19	DD--RY---DR-F-SF-----\$		493
HPV25	DDA-RY---DR-F-SF-----\$		502
HPV20	DE---Y---DR-F-SF-----\$		497
HPV21	DK---Y---DR-Y-SF-----\$		503
HPV14d	DQ---Y---DR-F-SF-----\$		483
HPV5	DEA-RY---DKAY-NL-----\$		514
HPV36	DDV-RY---EK-Y-NL-----\$		509
HPV47	DDA--Y---EW-Y-SL-----\$		506
HPV12	DE---Y---ETAY-NL-----\$		494
HPV8	DE---Y---DT-Y-NL-----\$		498
HPV24	-DL-RF---DW---SF----K-\$		467
HPV15	IKIM-L-P--DW---L----D-\$		456
HPV17	IQKM-L-P--DW---HL----D-\$		452
HPV37	-K-M-L-P--DW---HL----E-\$		454
HPV9	IR-M-L-PT-DW---NV----D-\$		461
HPV22	-QQM-L-L--EW-Y-QF----D-\$		436
HPV23	IQEM-L-L--DW-Y-QF----D-\$		431
HPV38	IQ-M-L-T--EW---QF----D-\$		441
HPV49	VKIM-L----EW-F-NF----K-\$		488
HPV4	VKHNLF--LCTTYT-SLN-----\$		402
HPV65	VKHNLF--LCTTYT-SLN-----\$		402
HPV48	---H---TTIT--RL-----\$		396
HPV50	---GL---NTTY-M-HL-----\$		396
HPV60	MKL-TL-R-CTYTF-TLN-----\$		404
BPV1	-KH-PL-P-MNI-GFTASL..DF\$		410
BPV2	-KH-PL-P-MNI-GFTASL..DF\$		411
EEPV	-KR-PL-P-MRAQALT-IA..DF\$		415
DPV	-QQ-PL-P-MSAHGVT-TV..DF\$		416
BPV4	-AS-RL---SAVK-AL----G-\$		408
HPV41	-KS----NIGLFRAHAE..K-\$		387
COPV	MDR-TF--S-R-FR-GL----E-\$		385
CRPV	-SR-PL-STT--F--NFY...G-\$		390
ROPV	-EE-P--RHM--FV-NFF...G-\$		266
HPV1a	-ER-AL-RS-S-F--QFN...GS\$		401
HPV63	-DK-VV--S-S-I--AF----GS\$		398
MnPV	-C----A-MT-EQCS-A...-V\$		542

**Appendix B: Secondary Structure Prediction from E2 Sequences**

Protein sequences such as the E2 sequences that display less than 30% similarity might nevertheless be shown to have similar structures. In general, we tend to learn more about structure from dissimilar (but homologous) proteins than from highly similar proteins.

This appendix summarizes the secondary structure predictions over the E2 HMM-predicted sequence as determined by several different algorithms, Gibrat, Levin, DPM, and SOPMA. Two consensus structures are also reported, one based on the four different algorithms, the other (at the top of the print-out) based on individual E4 sequences as analyzed by the SOPMA method. The derivation of an HMM model sequence ('most likely sequence') is discussed in appendix A and elsewhere in Part III (Farmer and Myers). The various methods for secondary structure prediction are also discussed elsewhere in Part III of this compendium.

The structural code encompasses lower and upper case letters for alpha-helix (h, H), beta-sheet (e, E), turns (t, T), and random coil (c, C). States that are predictable with greater confidence are shown in upper case. The criteria for designating a state as upper or lower case are spelled out in the general discussion of this approach in Part III: states that are predicted in upper case letters have i) scores that are equal to or greater than the median average for scores assigned to that state over all positions and ii) scores that are in the upper quartile of difference from the second highest predicted state. Hence the absolute and the relative scores must meet stringent requirements to warrant upper case prediction.

## E2 Appendix B

hpv_E2.allseqs.SOPM	h.....hhhhhhhHhHHHHHHhHh.	hcchhhhhhhhhhhhhhhhhheeeccccccccc	58
Gibrat_ALL_E2	-.....-	-HH-----H-----HHHHHHHHHE--CC--	58
Levin_ALL_E2	-.....-	.HHT----C---CCH-T-E--CCHTS---CCCTS--	58
DPM_ALL_E2	C.....C-----	CTTCC---E--E-E-H--EE---HHHHH-E-C-CEE	58
SOPMA_ALL_E2	C.....CC-----	.HH-----H---HHHHHHH-EE-CCEEE	58
Consensus_ALL_E2	C.....C-----	.HH-----H---HHHHHHH-E-CCC--	58
HPV54	-.....-	C.C-TT--T-----H---HHHHHHHT-CHHHHH-	58
HPV32	-.....-	.-----ETE-EE--E-----TTTHH-CCCTTE	58
HPV42	-.....-	.-----E-EE-EE---C---TECCC-EEE	58
HPV3	-.....T-CCCC--C-----	C.C-CCCCCCCCEEEEEE---HHHC-HH---C-CEE	58
HPV28	-.....CTT-C-----	C-T-CTC---T-H-T-T-HH-----TEEE	58
HPV10	-.....-	.HH-----H---HHH-T---E---TTEE	58
HPV29	-.....C-----	C.CT-----H-TT---TT-TTEC-TTEE	58
HPV61	EH.....H-----	C---CCC---EEETEETCCT-----TTTT---T-E-	60
HPV2a	-.....CC-C-----	EC.C-CCCCCCCCC-----C-----C--	58
HPV27	-.....-	.H-----H---HH-CTT---T---T--	58
HPV57	-.....-	.TT-----C-----H---HH-C-H-E---C-EE	58
HPV26	-.....CC---CCCCT-----	EE.C---TCCCCCEEECC---H---HHHHH---CCCC-H-	58
HPV51	-.....C---CCC---EEEEEE-----	TT---EEETTETCCEE---H-TH---H---TE-	58
HPV30	-.....-	C.-TTTCTTEEEEEEETCC-TCEE---C-----HHHHTHH	58
HPV53	-.....-	.T-CCCTTC-----ETTEE-----TT---HHC---HH	58
HPV56	-.....-	CCT-C-----TTCEE-----TT-----HEH	59
HPV66	-.....-	C-----EE---CTTEEE---C---TT-----EEE	58
HPV18	-CC..CH-----	THH---T-CETEEEEEEET---HH-TT-E---EEE	62
HPV45	-HCCCCH-----	T-TTTCTTEEEEEE---EH---E-H-H---C---EEE	64
HPV39	-HH.HHH-----	T.C---H---T---CCTTEEEETC-H---HHTTT-E---TTT-	63
HPV70	-HH.HHH-----	.H-----T-CCCTTEEEEEE---H-TTTT-----EEE	63
HPV59	-HH..HH-----	C-----T-E-----EEEEECC-----H-T---CC---E	62
HPV7	-.....CC-----	HH---CTT-EETEEEEEE-EE---E---EEEHHHTEE-	58
HPV40	-.....-	HH---T-EE-TEEEEC-E---EEETEE---EEE	58
HPV16	-.....CCC-C-----	T---TCETECCEETECH---HHHHHHHT-HT---TEE	58
HPV35h	E.....H-TTCCCCCCC-----	T.T---CECTTCCC-ETEEH---HHHHHHHTT-C---T-	59
HPV31	-.....CCC-C-----	C-H---TT-EE---TT-E---E---EEETT-CCC-TEE	58
HPV52	-.....C-----	H-T-----H-----TT---TE---TTEE	58
HPV33	-.....-	.H-CCCC---CEEEEEE-HH-C-TT---CTCHHH-	58
HPV58	-.....EE-----	TTTTCCC---EEE---THHC---TTTEHHHHHH-	58
RhPV1	-.....H-----	HH-----H---H-TTTT---CCCCHEE	59
HPV6b	-.....-	.H-----CC-----C-E---TH---T-CCCEE-	58
HPV11	-.....-	.HH-----C-H---HHHTT---TEE---EEE	58
HPV44	-.....CCCC-----	T---CEEETCC-C-E-EHHHHHH-HT---CHCTHEE	58
HPV55	-.....-	TTCEE---C-C-EHHHHHHHHHHHHH-C-C-EEE	58
HPV13	-.....-	HH-----EEEEEHHHTE---TTECC-TEE	58
PCPV1	-.....-	T---TTCC-----HH---T-CC---EE-	58
HPV34	-.....H-----	T.T-H-E---TTCC---H-EHHHHHTHHETEC-CCTTE	59
HPV19	C.....CC-CCCTCC-----	HH---CCCEECC---E---H-----C--	58
HPV25	C.....CC-TTCC-----	CCCC.CHH---CCCCCCCC-TC-E---H---E-CC---	58
HPV20	-.....C-C-----	C-C-----C---E-----E---E--	58
HPV21	C.....CC-TCCC-----	HH-----TCCCC-T-----C---EEE	58
HPV14d	-.....CC---C-----	C-H---C-CCC-C---E-----C-CCEE	58
HPV5	-.....-	HH-----CC-CCCE-----E---E--	58
HPV36	-.....C-----	HH-----CC-----H-----E--	58
HPV47	-.....C-----	HH-----E-CCC-TCCE-----C-C--	58
HPV12	C.....CC---C-----	C-CCC--CC-TCECC-C-E-----CC--	58
HPV8	-.....-	HH-----EEC-T---E-----TE---C--	58
HPV24	-.....C-TTCCCCC---EEECCC-----	C---C---C---E-----E--	58
HPV15	-.....C-----	T---CCCCEEEEETTT-TTEE-----ETTTTE-CCT-	58
HPV17	-.....CC-TCCC-----	C-CC-EEECTTCTC-C-EE---TT---E---E--	58
HPV37	-.....CC---C-----	TCC---TTE---CTE---TTT---E--	58
HPV9	-.....-	TTTCCCC-H---HHHH---T-E-----	58
HPV22	-.....-	CHH-----H---H-TT-TTE---CC--	58
HPV23	-.....-	.HH-----C-E---TTT---EC-	58
HPV38	C.....CE-TTCEEEE---EEECC-----	C---TCCCCCTCECEE---HHHH---TE-----	58
HPV49	-.....TCCCCCEE---EEECC-----	C---CC-EC-TCCC---E---E--	58
HPV4	-.....-	HCC-----TC---E---HHHH---HTTT-	58
HPV65	-.....EEEEECC.TTCC-----	CTC---T---T---HHHHH---CHCC--	58
HPV48	CCT.TCC---C-C---EEECCC-----	C-TTCCTTTEEEE-----HHHHHTTTEE---C--	63
HPV50	CCHHHHH-----	EEEC.C---TCCTCEE---H---HHHHHTTTE---E--	64
HPV60	CC....CCECC-----	HTC-----H-EEE---ETTT-C--C--	60
BPV1	-.....T-----	E-CCCC---EEEEECC---H---TT-TEE---T-	58
BPV2	-.....T-----	HT-T---EEEEECC---HHHHHH-TEE-----	58
EEPV	EC....C-CCCC---CCCC-----	C.C---CCCTCEE-----TT-E---C-----C-CEE	60
DPV	C.....CC-----	CCCEECC---TCCCTTECCCCC-T-TTH-HH---EE-----	58
BPV4	-.....	EEEC.TTTTC-EEEEEECC---TTCC---EE-TTTTT-EEE	58
HPV41	-H...HH-----	THHECTT---EEE---CCCC---EETTT-CCTTH-	61
COPV	-.....C-----	EE.T-C-----CTCCEEC-TTCE---E-T-E---E--	58
CRPV	-.....-	H---C-----T-CCTE---E-TT-C-TT---	58
HPV1a	-.....C-----	C-H---CCCCCT-C-CC---H---T-E---EE-	58
HPV63	-.....-	CHH-C---EEECC---H---TTT-E---EEE	58
MnPV	-.....-	HH-----CC-TTTE---CH---CCCC--	58



## E2 Appendix B

hpv_E2.allseqs.SOPM	ccccccceeeeeeeeeeccc.t.cc	checcccccc.tteeee..etccceeeeeecTtccccT...	180
Gibrat_ALL_E2	---HHHH-H---C..C.-H-E-E-E.EE-----.	-C-----HHHHHHHHHHHHC..	179
Levin_ALL_E2	--T-HHHHHCHH--CC--..SSECCTT--EE.EE-----.	T-SH-HS---HH-TTHTC..	179
DPM_ALL_E2	T---HHH-----CC..C.--CC-E-TE--.CC-----.	CC--T-----HC-HT-----.	179
SOPMA_ALL_E2	---HHHHHHHHHHH--CC..C.-H-HC-----.	CCCHHC..CC---C---CHHHHHHHC..	179
Consensus_ALL_E2	---HHHHH-HH--CC..C.--CCE-----.	CC-----CC-----HHHHHHHC..	179
HPV54	T-----C-----C-T..C-E-HHHHHHHH.HH-----.	-THE-----C-HH---C..	181
HPV32	-----CCC----ET..TTCC-EEEEEEE..-C-----.	HC-HH-H-CCT-C-TET..	180
HPV42	---EE--T-----E..-T-E-EEEE-E.HHH--H..	HHHHHHHHHHC--HHT-T..	180
HPV3	----HH-----CT-CC..-H-EEE--T.CC-----.	C-T-----HT-C-HHEE..	181
HPV28	---EEE--C-CC--CT-CC.T-E-EEEE-E-----.	.ET-T-----C-EEEC..	181
HPV10	---HEE--CCHH--CTTT..TTE-E-TEEE..CCC..	H-T-C---TT-C-EEEC..	181
HPV29	HHHEE-HHCCCC--TC..T-H-ETTEEE..---C..	EEEET--HT---EEC..	181
HPV61	-TT-TE---C-C---C-T..HEEE-----EEH..-----	TTTT-----EHHTT..	183
HPV2a	----E---TTT--TC--CE..-CCC-----EE..-----	E-----C-CC--EEE..	181
HPV27	--TTE--CC--HTH..TTCH--HHH.HE-----.	T-T-E-----C-C-TEEE..	181
HPV57	---EE--CT---TT-CC..TCCC--E..CE-----.	T-E-----TCCTTEE..	181
HPV26	----C-CC-C-CHCC..C.-H-CTT-T..CC--C..C--	TT---C---HHH-C..	179
HPV51	---TE--TTTT---T..T-E-E..E.CE-----.	.T-T-----EEE..	179
HPV30	T-----C..TTTC-EEEEEEE..E-----.	CTTE---CCC--TEEE..	179
HPV53	T---HH--HHHHHH-C..C.--EEE..CCC-----.	C---T---CCC-----C..	179
HPV56	---H-C-----C..-TTCCCHHTT..---.	C-E-----TT..	180
HPV66	T---HH-----C..-T-EC-EEE--EE..E---C..	CC--E---CCH..-----T..	179
HPV18	--TTEE--CCCC--T..C---TC--HHH..H..-----.	T-----T-CHHH..	184
HPV45	--T-EE--CCTC-----THE--HHHHHHEE..E---C..	C-----T-HH-T..	186
HPV39	-----CCHHH-C..CC-EE-CTT--HTHC..E---C..	HHHHHTTH--EHCHT..	185
HPV70	-----T..-TTE--TTHEEHTH..H--C..HHHHH..HHTT--T-TC..	..	185
HPV59	-T--E---CCT--CC..HTT-E-----T-EET..-----.	TC--T-----E-CH--C..	184
HPV7	--T-E-----TH..-T-HHHHT-HHH.CCC..CC---C--	T-C--T..	179
HPV40	--TT-E-----T..H.HH-HHH..H..CCC..-C---T..	E-H-TTT..	179
HPV16	----HHH-TCCT-----EE--E-T--EE..EE..-----	H-----HH..HTT..	179
HPV35h	---T--CCCC-----H..C..-HHHHHHT..-C..-----	T-----HH..HTT..	180
HPV31	T-----CCTCC-----T..-T-E-EE-----T-C..	..-T-E-----HH..HTTC..	179
HPV52	TTT---CCTCCC-----T..TTCC--EE..EE..C..-----	T-----TTEE..	179
HPV33	--T--CC-TTTC--E..---C--EE--EE..-----.	CC-TE-----HT--HTT..	179
HPV58	T--EE--CCTT--E..---C--EEH--EEE..EE..-----	C-T--T--HHH..	179
RhPV1	-TT--T-----T..C..-H-HHHHE-T..TC..E..-----	ET-T-----EHHEETCCTT	183
HPV6b	-TT--T---CCC---T..-T--T-EHHHH..-CC..C..	-C-----C--HHE..HHHH-C..	179
HPV11	--HHHH-----T..-TTE--E-EHHHH..H-T-E-----	C-----ETC..	179
HPV44	--HHHHHHH-T-----TTE--EEEEE..---C-----	CC--HHH..	179
HPV55	---HHHTCH--T..-T-E--E--EEH..---C-----C--	CC--T-T..	179
HPV13	----T-TCC-----E..-TTCTTHEEEEHH..CC..C..	TC-----CHHHHHHH-C..	179
PCPV1	--T--C-----CT..C..-HHHHHHHH..HHH..HCT--C--	CT-H-H-C..	179
HPV34	TTT-----E..-TTE--EE--H..H..HH..C..HC--H..	HT--CHHHT..	181
HPV19	----C-----HH..C..CCC-----CE..E..H..T-----	E-C--H..	180
HPV25	-T---CCC-CC--H-----HCCC-----.	C-THEHH..E-----	180
HPV20	TT---HH-----C..-T-E-----CE..-----.	CE-----E-CT--C..	180
HPV21	-T-----CCCCC-CC-----ECC-----CE..C..EEET..	ECC-----	180
HPV14d	-T---CC--T-CC-C..-TH..HH..CE..-----.	TH--HCCH..	180
HPV5	TT-----HHHHHHH..-TH-HC--TT-EE..EE..-----.	T-THHTH..E--TT-T..	180
HPV36	-----CC--C-----T..-T-EHC-----CE..C..TEE..	..CHHC..	179
HPV47	-----CC-----C-T..-E-C..-----.	T-----E-C--H..	180
HPV12	-----C-----C..-ECC-----T..CC--CC..C-T..	C-C..	180
HPV8	-----HH-----CCCC-T..-EE..T..E--C..T..TEET..	E-CHHHH..	180
HPV24	-----C-----C..C..CCC-----C..-----.	H--TH--ECC--C..	180
HPV15	-T--EE--TT--H..H..HH..HH..E..-C..-C..	-C-HHHHHHHHHTCCHHHTC..	180
HPV17	--HHH..HH-----C-T..-T-HC..---C..-E..EE..	---ECCTTTTC..	180
HPV37	--TTEEE-----H..H..H..CC--HHH..CC..C..EE..	---CT-CTTE..	180
HPV9	-T---E--C--C--H..-H..HH..HHTTT..TTE..EE..C..	CHHHE-----CCHHHT..	180
HPV22	-T-E-----T..-T-H-HEE..HHT..-----.	EEEEE..E--TTTT..	180
HPV23	--T--C-----CCC--C..-H..HHHHHHHH..-C..C..	HHHHHTHH..H-----E..	180
HPV38	-----C-C-----C-T..-HHH..-C..TTCC..C..EE..	CC-CC-----	180
HPV49	-----EE--CCTCCC-TC..T..CCET--EE..H..CCHT..C..	C-H--T-----ECC-EET..	180
HPV4	--TEEE--CCC-----C-T..H..HH..HHHH..---C..CC..	CC-----E--TTHEE..	181
HPV65	TT--TE-CCTCHHHHHHH..HH..HHTTT..HE..-----.	C..CC--T-----C--T..	181
HPV48	--TT-CT-----CT..-T-TCCC--EEE..T..CC..E..	CC--E-----C-C--E..	186
HPV50	--T--C-----C--C..-T-H-TCHHTEE..CCC..T..CC..	-----T-C-TEEE..	187
HPV60	--TT--CCCC-----CT..-HHT..HC--T-E..CC..C..	CC--C-----E..-T--C..	183
BPV1	-----C-----TH..HCH..C..HE--E..TT..CE..E..	EE--TE-----CCTCCT..EEE..	181
BPV2	-----C-----HCCT..C..-E..TTTH..CHH..H..HHH..	TE-----CTT-CHHEEE..	181
EEPV	-----CC..-HHHH..TT..CC..-T..TC..-C..CC--	CC--CC-----	184
DPV	TT-----CC--C..CC..HE..-----TT..C..CC..T..	CC--T-----HHCH..T..	182
BPV4	TT--HE--HHHTH..EE..E..TCCC--TT..C..C..T..	T-----HHH-C-HHHH..	181
HPV41	--TT--HHH..HHHHHTC..T..-E..TT..T..EC..E..	T..E-----CC--TT--C..	184
COPV	--T--HE--C..HH..CH..H..HCTT..CE..C..C..T..	E-----HHC..	180
CRPV	TTT---CC-----TT..-T-E--E..TT..E..C..TTT..	-----CC-T--C..	181
ROPV	-----CC--C-HC--TTT..-H..HHHHHHHH..EE..C..	TCC--C..	34
HPV1a	-----TCCC-----CTT..C..-THCHTTHHH..E..C..T..	-----T..ECC--C..	181
HPV63	-----EE--CTC--C..-CCC..-C..CC..C..C..C..	E-C-T-EE..	181
MnPV	-----		180

hpv_E2.allseqs.SOPM	...ccceeeeettccee...ee..cCCCcCC.....	CcCCCCcc...cccC..cccc	218
Gibrat_ALL_E2	.....EC-E-----E-----.	EE-EEEE...EEEE..EEEE	215
Levin_ALL_E2	....TTTCCH-E--E-----.	.....-E-S..---E	215
DPM_ALL_E2	....T--C---EC-E-----E-----.	----HHE...E-----	215
SOPMA_ALL_E2	....C---EC-E-----E-----.	----EEE...EEE-----	215
Consensus_ALL_E2	....C---EC-E-----E-----.	----EEE...EE-----	215
HPV54	....T-TC---E-TE-----E-----.	HHHHHHEEE...TT-----	218
HPV32	....EEE---CT-C-E-----EE-----.	TEE---EH...HTT...-HHH	218
HPV42	....-E---CCC-E-----EET-----.	-----.-.----E...---T	213
HPV3	....ETTC---E-----C-----.	-----.-.----E-----	217
HPV28	....T-C---E-TE-----T...EEE-TE-----.	EEE..EE...TT-----E--	217
HPV10	....-CHH-HE-TE-----T...EEE-----.	---E..E-----	217
HPV29	....E-T---E-T-----T...T...EE-----.	EHH...-----T-----	217
HPV61	....ET-----EEE-E...CCT-----HH-----.	HHH..HHH...HHHH..HHTT-	222
HPV2a	....E-E---EE-E-----C-----E-----.	-----.-.----E-----	217
HPV27	....-EE---EETE-----H...HEE-----.	-TT..E...HHH...T--H	217
HPV57	....EEE---C-E-----EH..HEH-----.	..HH..HHH...HT--..EE-E	217
HPV26	....-E---E-T-C-----T-E-----.	-----C-----	216
HPV51	....T-H---E-T-CC-----H-----.	-----.-.----EHHH	214
HPV30	....ETTC---EC-C-----E-T-----.	EE-----H..HEEH	215
HPV53	....TE-C---E-T-C-----E-T-----.	EE-----HHHH..HHHH	215
HPV56	....EEE---ECT-C-----E-TT-H-----.	---E-----HH..HHHH	216
HPV66	....EEECTTHH-HHH-----EETT-----.	-----EE-----T-----HHH	215
HPV18	....T--C-----CCC.T-----.	---H...-----HHH..HHHH	220
HPV45	....-C---T-C...CCC-----.	---H..H...HHH..HHHH	222
HPV39	....TT-H---T-----CCC..EEE-----.	-----.-.----EEE..EEE	221
HPV70	....-TE-----CTT---CC..EEE-----.	-----.-.----HHH..HHHH	221
HPV59	....EHHH---T-CTTHH...HHH..HT-----.	....E...H---.---TT	220
HPV7	....-CC---E---C-----T...TTE-----.	....EE-----	213
HPV40	....CC---E---E-----T...THE-----.	H-T..EEE...T-CC-----	218
HPV16	....T-E---E-----E-T..EE-----.	-----.-.----H..HEHH	213
HPV35h	....-E---E-T-----E..E-----.	..HH..-----HHH..HHHH	214
HPV31	....TT-----C-TE-----EEETEEE-----.	-----.-.----HEE..EEE	213
HPV52	....EEE---E-TE-----E-----.	-----EE..EEE	212
HPV33	....T-HHH---T-----E-----.	-----.-.----	212
HPV58	....T-E-----T-----.	T-----.-.----T	212
RhPV1	CCC---E---HHH-THH...C..HT-----.	HHTHT-C...EEE..EEEH	223
HPV6b	....T-----C-E-TE-----T-----.	-----EE-----E	214
HPV11	....-T-C---CE-TE-----TT-----.	-----EE..EE---T--E	214
HPV44	....H---E-T-----TT-----.	-----H..ETT-----	214
HPV55	....T-HHHCC-E-TT-C-----.	..HH..E..E-----	214
HPV13	....-EE---C-E-E-----T-----.	---E..HEE..EEE-----E	214
PCPV1	....TTTTC---E-T-C-----TTT-----.	..HH..EEE..EE-----	214
HPV34	....ETT---C-T-----EE..EEE-----.	....T-----T..EEE	214
HPV19	....-CC---C-E..ECCC-----.	-----C-----CC-----	224
HPV25	....T---CCCCC---C-E..E-CC-----.	-----C-----CC-----	224
HPV20	....-CC-C---C-E..E-CC-----.	-----T-----C-----T-CC-EE-	224
HPV21	....-CCCCC---CCE..CCCC-----.	-----C-----CC-----	224
HPV14d	....T--CC---C-E..E-CC-----.	-----C-----CC-----	224
HPV5	....-CCHHH---E..E-C-----T-----.	-----C-----CC-----E	224
HPV36	....T---CCCCC---C-E..E-CC-----.	-----C-----CC-EE	223
HPV47	....T--CC---C-E..ECCC-----.	-----C-----CC-EE	224
HPV12	....T--CC-C---T-C-E..ECCC-----.	-----CC-----	223
HPV8	....T-TCC---C-E..E-C-----T-----.	-----C-----CC-----	224
HPV24	....-CC---C-E..E-CC-----.	-----.	221
HPV15	....TTT---ECTTC-E..CC-C-----.	-----HC...-----CC-----	223
HPV17	....-HC---EC-T-E..EC-CC-----T-----.	T-----C-----CC-----	225
HPV37	....-TTC---EC-T-E..CC-EC-----.	-----HHEEE..EEE-CCE-EE	226
HPV9	....-TE---C-----E..E-C-----.	-----EEE..EEET-CC---T	226
HPV22	....-TT---E-T-E..ECCC-----EE-----.	E-----T..T-----.	219
HPV23	....TEE---ECT-C-E..ECCC-----.	-----.	221
HPV38	....-E-C-CC---C-E..ECCC-----.	-----.-.----H-CC-----	222
HPV49	....-CC---EC-C-E..EC-C-----CCCCCCCCCCCC-----.	-----CC-EE	234
HPV4	....HH---EEH-CC-----EEE-----.	-----TT-CC-----	218
HPV65	....-TT---EE-E-----EE..EE-----.	-----EE-CC-----	218
HPV48	....T-E---E---C-----EE-----.	-----C-----TH..HEEE	220
HPV50	....ETTC-C---E-T-CC-----T..T-----.	-----HHH..T-CC-----	224
HPV60	....-E---E-----E-----.	-----C..E---.TT-	219
BPV1	....E--C---T-----E-----.	-----TT-EE..EE---TC---	223
BPV2	....E--CCCT---E-H..H-----.	-----E..EET-CC-----	222
EEPV	....T--C-CCC---C-E..E-E-----.	-----EEE...TT..-----	227
DPV	....EEE---C-TC-----T...HHHHH..EEECTT...T-----C-----	-----C-----	227
BPV4	....TT-CCH---ECHE-C-----THEEE-----.	EEE..EE...-----CTT--T	216
HPV41	....ET---H-TE-----T-E-----.	EE...-----EEE	218
COPV	....-CC---CC...CCC-----.	-----TT..---T..T-E	218
CRPV	....-TTCC---ECH-CC-----.	-----E..-----T..C---T	216
ROPV	....-T-CC---E-TCC-----EEE-----.	-----EEC.C---CC-----	75
HPV1a	....-TT---E-T-C..E-E-----.	-----T---EEE..TT---.T---	222
HPV63	....TT-C---E-T-----E..E..T-----.	-----C..-----CC-T-	222
MnPV	....EEE---CCCCC---EEE-CC-----.	-----EE..E-----	224

## E2 Appendix B

hpv_E2.allseqs.SOPM	ccc.....	CCC.CCCC	229
Gibrat_ALL_E2	EE.....	EEEEEE	222
Levin_ALL_E2	E-.....	EE-EE	222
DPM_ALL_E2	E-.....	--	222
SOPMA_ALL_E2	--.....	EEEEEE	222
Consensus_ALL_E2	E-.....	EE-EE	222
HPV54	--.....	--	228
HPV32	HHH.....	HE-C	230
HPV42	--.....	HHHHHHH-E	228
HPV3	.E.....	--	222
HPV28	T-.....	--H-	223
HPV10	--.....	--TT	223
HPV29	-T.....	--	224
HPV61	--.....	T-T--	230
HPV2a	--.....	--C	229
HPV27	EEE.....	EE-C	229
HPV57	HHH.....	HHHHHHHHH	229
HPV26	--.....	--	225
HPV51	ET.....	--	221
HPV30	HT.....	C-TTE-	224
HPV53	HH.....	-T--	223
HPV56	--.....	--EE	224
HPV66	--.....	--EE	223
HPV18	H-.....	--	227
HPV45	HH.....	T--	229
HPV39	E-.....	--EE	228
HPV70	H-.....	--	228
HPV59	--.....	-T--T-	230
HPV7	-H.....	--HHH	220
HPV40	--.....	--	228
HPV16	.HH.....	HH--HH-	222
HPV35h	HH.....	HC-TT-	223
HPV31	--.....	--	223
HPV52	E-.....	--	218
HPV33	--.....	--T.	222
HPV58	.T.....	--	215
RhPV1	EE.....	EE-T-	230
HPV6b	--.....	--TT	219
HPV11	--.....	--	220
HPV44	--.....	-T-EEEEEH	225
HPV55	--.....	--EEHHH	225
HPV13	.EE.....	--E-EE	221
PCPV1	--.....	.EEHHH	221
HPV34	E-.....	--	221
HPV19	--CCCCCCCCCCCC	CCCCCCCECCC--C	261
HPV25	--.ECCCCCCCCCCC	CCCCCCCEEECCC--T.	258
HPV20	--CCCCCCCCCCC.	CCCCCCETTC--TC	258
HPV21	--CCCCCCCCCCCC	CCCCCCCTCCC--TC	259
HPV14d	--CCCCCCCCCCCC	CCCCCCCEEECCC--C	261
HPV5	EEEECCCCHEEEETCCC	CCCCCTCCCTCC-TT	261
HPV36	.EECCCCCCCCCCCCC	CCCCCCCTCCCCCCC--	260
HPV47	--CCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC--EEE	--	288
HPV12	--CCCCCCCCCCC.	CCCCCCCTCCCC--	255
HPV8	--CCCCCCCTCCTT	CCCCCTCETCCC--T	259
HPV24	--CCCCCCCCCCCCCCCC	CCCCCCCTCCCCCCCCT--	266
HPV15	H-HCCCCCTTCCCC	CCCCCCCCCTT--CEE-E	256
HPV17	--TCHCCCCCCC	CCCTCCCTEE--TT	255
HPV37	E--CCECCTCCC	CTCCEEEE--	254
HPV9	--EEEEETTCCC	CCCCEEEETT--	255
HPV22	--CCCCCCC	EEETCTT--CEESEE	245
HPV23	--TTCCCCCCC	CEETCEE--TEHEH	245
HPV38	--CCCCCCCC	CCCECCCC--	248
HPV49	EE--CCCCCCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCCCCCC--EC	285
HPV4	.TT.....	--CHHHHH	229
HPV65	--.....	--EEE	227
HPV48	E-.....	--	228
HPV50	-E.....	--	231
HPV60	-E.....	EEECT--	233
BPV1	-T.....	--	233
BPV2	TT.....	--H-	232
EEPV	.-.....	TCCCCC--	242
DPV	--.....	EEE.E--T	237
BPV4	T-.....	CTTT--TT	229
HPV41	E-.....	--	225
COPV	.E.....	EE--	224
CRPV	--.....	--C--HH-	227
ROPV	--.....	--EE.	85
HPV1a	--CCCCC	EEEE--	242
HPV63	--.....	HHH.HHH--	233
MnPV	--CCCCCCCCCEE	CCCCCCCCCTCCC--EEE	259

hpv_E2.allseqs.SOPM	CCcc...cc...CCCCCC.CCCCC...CCCcccC.....ccCC.....	257
Gibrat_ALL_E2	EEEE...EE.....-EE-EE..EEEE----...--E.....	252
Levin_ALL_E2	-----...-----C..C.....	252
DPM_ALL_E2	--HH...-T...-T-T...-----C..TTTTT----...-T.....	252
SOPMA_ALL_E2	E-----...-----C..C.....	252
Consensus_ALL_E2	-----...-----C..C.....	252
HPV54	-----...-----	244
HPV32	---E...-----T--EEEEE-----T-----E.....	261
HPV42	EEE...EECCC--HH---EE...E-E---...-----H.....	259
HPV3	-----...-----C...C-----H-----	251
HPV28	--H...HH...E--T...-----T...HHHHHHHH...HHTT.....	250
HPV10	-----HH...HHHH...HHH...HHHHHHH...HTHT.....	249
HPV29	-----...-----HHHH...HEEEECCCCCCCC-----	258
HPV61	--EE...EEC-----CEE-----...-----T--H.....	253
HPV2a	---C..C-----C--E...C-----	261
HPV27	--H...-----EEE...C-----	258
HPV57	HHH...E-----HH...H-----	253
HPV26	--E...E-----	244
HPV51	---EE...-----HHHHHTH-----H.....	240
HPV30	TT-----E...E-----HHHH-----TT-----	245
HPV53	-----E...EEEE-----	242
HPV56	EE-----TT--HH..HEH-----TT-----	244
HPV66	EEE...EE...E--HH..HHT-----T-----	243
HPV18	---E...EE-----E..E-----T-----	248
HPV45	-TT-----	250
HPV39	-----...-----	248
HPV70	--HH-----T-----E-----	248
HPV59	-EE...-----T...T-C-TTT-----	248
HPV7	-HEH...-----HH-H..HH-E...EE---T--	249
HPV40	-----E-----	243
HPV16	HHEE...EE...-----H...EE-----	240
HPV35h	--E...EE...ET-----	242
HPV31	-----E-EE--C--E-----	245
HPV52	--E...EE...EE-----	238
HPV33	HHHH-----	243
HPV58	HHH...-----HH-----H-----	228
RhPV1	-----H..HH-----HH-----	249
HPV6b	-EE...E-----T-----H-----	239
HPV11	--E...E-----T..T-----	240
HPV44	TTT-----HH..HH-----	246
HPV55	--TT...-----T-----	246
HPV13	-EEE...EE...E-HHH..T-----	242
PCPV1	HHHH...H-----HHHH...T-----	242
HPV34	-HHH...HT-----	235
HPV19	-----C..C--E--C-----CCCCCCCCCC.....	301
HPV25	-----C-----C..C-----CCCCCCCCCC.....	308
HPV20	-----C-----C..C-----H-----CCCCCCCCCEEECCCCC.....	306
HPV21	-----C..C-----C..C-----CCEEECCCCCCCCCEEE.....	306
HPV14d	-----C..C-----E-----CCCCCCCCECCCC.....	303
HPV5	-----T..T-----C..C-----TT-EEEE..CCC--T-CCEEEEEECCCC.....	312
HPV36	-----C..C-----C-----CCCCCCCCCTCCCCEEE.....	307
HPV47	EE-----C-----C..C-----E-----E-E-----	318
HPV12	-----C-----C-----E-----EEECEECCCCCCCCEEE.....	299
HPV8	-----C..C-----EEE-----CCHEEECCCCCCCCCTT	311
HPV24	-----C-----	293
HPV15	-H-T...TT...T-HHH..HHHHH...C-H-TT--HE...EET-----	288
HPV17	-----C..C-----TTT..C-----EEE-----	283
HPV37	-H-T...TT...T..T..T..HEEEHC..CE--T--CHECCCC-----	288
HPV9	-----C..C-----CT-HHHH..C-----CCCCTCCC-----	294
HPV22	EHHH...TTT..TT-T-H..HH-H...HEHHH-H.....TTTT-----	276
HPV23	HHTT...TT...T..T..HHEEHHTHH..HHHHHH-H.....TTTT-----	273
HPV38	-----T..T-----C-----C-----	277
HPV49	-----C..C-----	314
HPV4	-----HH..-----C-----C--TE-E.....EEE-----	258
HPV65	E-----T..-----CC..CTT--EE.....EEE-----	258
HPV48	-TT-----C-----C--HH.....HT-----	256
HPV50	--T-----T-H-----EEE-----E-TT-----	259
HPV60	T-----C-----C--HHEE.....HHH-----	258
BPV1	-HEE...ET...T---EEEEE...E-TTTT-----EEE-----	264
BPV2	HHHE...E-----E-EEEEE..E-----EEE-----	263
EEPV	---CCC-----E..C-----	273
DPV	---C..EE...HH-----C..C-----	266
BPV4	-----TTT-----C..C-----T-----	259
HPV41	-----HHH-----C-----T-----	243
COPV	--T...-----TT--T..T-E-----TT-----	244
CRPV	-----...-----EEEEE...-----T--TT-----	251
ROPV	-----C-----C--T--CCE..CC-----	118
HPV1a	--E-----T...-----T..T-----TT-----	272
HPV63	T-----HC---TT..T-----E-----	263
MnPV	-----E-----C--H---C.....C---CCCCCECCCCCCCCCTC	309

## E2 Appendix B

hpv_E2.allseqs.SOPM .....	257
Gibrat_ALL_E2 .....	252
Levin_ALL_E2 .....	252
DPM_ALL_E2 .....	252
SOPMA_ALL_E2 .....	252
Consensus_ALL_E2 .....	252
HPV54 .....	244
HPV32 .....	261
HPV42 .....	259
HPV3 .....	251
HPV28 .....	250
HPV10 .....	249
HPV29 .....	258
HPV61 .....	253
HPV2a .....	261
HPV27 .....	258
HPV57 .....	253
HPV26 .....	244
HPV51 .....	240
HPV30 .....	245
HPV53 .....	242
HPV56 .....	244
HPV66 .....	243
HPV18 .....	248
HPV45 .....	250
HPV39 .....	248
HPV70 .....	248
HPV59 .....	248
HPV7 .....	249
HPV40 .....	243
HPV16 .....	240
HPV35h .....	242
HPV31 .....	245
HPV52 .....	238
HPV33 .....	243
HPV58 .....	228
RhPV1 .....	249
HPV6b .....	239
HPV11 .....	240
HPV44 .....	246
HPV55 .....	246
HPV13 .....	242
PCPV1 .....	242
HPV34 .....	235
HPV19 .....	301
HPV25 .....	CCCCCCC 315
HPV20 .....	.....CEEEE 311
HPV21 .....	.....CEEEEEECCCCC 323
HPV14d .....	.....C 304
HPV5 .....	.....TCCC 316
HPV36 .....	.....ECCCC 312
HPV47 .....	..... 318
HPV12 .....	.....CCCCCCCCCCC 313
HPV8 .....	.....CEEEEE 317
HPV24 .....	..... 293
HPV15 .....	..... 288
HPV17 .....	..... 283
HPV37 .....	..... 288
HPV9 .....	..... 294
HPV22 .....	..... 276
HPV23 .....	..... 273
HPV38 .....	..... 277
HPV49 .....	..... 314
HPV4 .....	..... 258
HPV65 .....	..... 258
HPV48 .....	..... 256
HPV50 .....	..... 259
HPV60 .....	..... 258
BPV1 .....	..... 264
BPV2 .....	..... 263
EEPV .....	..... 273
DPV .....	..... 266
BPV4 .....	..... 259
HPV41 .....	..... 243
COPV .....	..... 244
CRPV .....	..... 251
ROPV .....	..... 118
HPV1a .....	..... 272
HPV63 .....	..... 263
MnPV .....	..... 374
CCCCCCCCCCCCCCCCCCCCCTCCCCCTCCCCCCCCCCCCCCCCCHHHHHHHHCCCEECCCT	

hpv_E2.allseqs.SOPM	.....ccCcc.CCC...cCCCCCcCcc.....	ccccCCcCcccc	287
Gibrat_ALL_E2	.....-EEEE--...EEE---EEEE.....	E---EEEEEEE	282
Levin_ALL_E2	.....S---C-S-.....	-T---E-TT--	282
DPM_ALL_E2	.....TTT-TCT-----TTTT.....	TT--T--TTTT	282
SOPMA_ALL_E2	.....----EEE...EE.....	----EE-----	282
Consensus_ALL_E2	.....----EC-.....	----EE-TT--	282
HPV54	.....-EE...EE.....T--..	.....-----T	266
HPV32	.....E-E.....EEE.....	TE-E---TT-T	288
HPV42	.....H-T-TC--T...EEEET--T-	C---E-EH---ET	291
HPV3	.....----C-----	-----	282
HPV28	.....T----E-----TTTEE--	-----T-EEEE-	275
HPV10	.....T---TT-T-----HHHH.....	HHHTT--H-HHHH	276
HPV29	.....-TT-----T--H.....	EE-T-----	288
HPV61	.....HHH...T-----T-TT-----	T---EE---EE	280
HPV2a	.....-EEEE-----E-----	--T---E---	289
HPV27	.....EEEE..T-----EE-----	E---E-E-T	286
HPV57	.....-EEE-----TT-T--TTE-	EET-----T--	281
HPV26	.....-----EEEEEEE-C-----	CC-----E-----	276
HPV51	.....-EE-----EE-----	EE...EE---T-	261
HPV30	.....-----H-----	HH...HHHHHHH	268
HPV53	.....T-----	HHHHHHHHHHHHH	271
HPV56	.....TH--T..T-----HHHHHH-HE	EE---E-----TT	275
HPV66	.....T--T..T-----EEE-----	E...EE-HHH-	269
HPV18	.....-----TT-----	-----EEEE	268
HPV45	.....-TT..EE-----	-----H--	270
HPV39	.....-TT-E..EE-----	T---EEHH--	270
HPV70	.....E-----	--HEEEE....	262
HPV59	.....TT---EE-----	EE...EEE-----	270
HPV7	.....-----EEEH-HTTT-----	T-E...EEE-----	274
HPV40	.....-T-----EEEET-----	-----TTT---THH	269
HPV16	.....-----TT--HHE-----	HHH-----	265
HPV35h	.....ETT...T-----E---T-----	EEEEEE-H-T	267
HPV31	.....-----EE-----	E-----ET	272
HPV52	.....-----TT-----H-----	E-----H-----	267
HPV33	.....EE-----	T.....-----	253
HPV58	.....TTE-----TT-E--TC-----	EEE-...TTT-TT	259
RhPV1	.....HH-TTT-----	EE...EEET--	268
HPV6b	.....--EE-C-T-----EE-----	EEE-----	264
HPV11	.....-----EE-----	EHHHHHHH-HHHH	263
HPV44	.....T-----TEEE-----	E-----	274
HPV55	.....T-----HEE-----	EE-H--H-----	274
HPV13	.....-C-----TT--EEEEEE-----	EEEE--HHH--	273
PCPV1	.....TT--T-T..T-----T-EEE-----	HHHHHHHHHHHTT-	273
HPV34	.....TT-----	-----HEEHTT	249
HPV19	.....CCCCCCCCC-C-T-----EE---TCCCCEC.....	CCCTTC-C-E-----	356
HPV25	.....CCCCCCCCCCC-T-----CCCC-----	CCTC--TT-T-----	364
HPV20	.....EEEEEECCCCC-TTC-----E---EEE-----	EETT--TTTT--TT-----	358
HPV21	.....CCCCCCCCCCC-----	T---EEE-----	363
HPV14d	.....CCCCCEE-----T-----	-----T-----	344
HPV5	.....TTEEECCCTTC-C-T-T-----EE---HHHHHHHHHEE.....	ETTCCCCC-TT-T-T-----	376
HPV36	.....CCCCCTCCCCC-C-----C-----	CCEEECCC...CCCCCCCCCT-----T-----	373
HPV47	.....T..T-----EECCCCCTCCCCCCCCCCCC-----T-----	368	
HPV12	.....CTTCCCCCCCC-----T-----	C-----	355
HPV8	.....CCCCCCCCCTTC-T-C-----EEE---TT-----	C-----T--T--	361
HPV24	.....E---C-----	-----E-----	323
HPV15	.....-----C-T-----EE---HE-TT-----	C-----	320
HPV17	.....-----EETT-----TT-TT-----	TT--TT-----	314
HPV37	.....E---TEEE-----	-----T--TT-----	318
HPV9	.....T---TT-----	TT-TT-----	324
HPV22	.....--HC-----HH-H--TT-T-----	T-----EE-----	308
HPV23	.....---EEC-TT...TT-E-----TTTT-----	EETT--EEE-----	305
HPV38	.....-----C-T-----T-----	-----T-----	307
HPV49	.....-----TC-EE...EEE-----	C-----TT-----	346
HPV4	.....TT-C-T-----E-ET-----	-----E--	289
HPV65	.....---TTC-----TT-T-----	C-----TT-EEEE-----	290
HPV48	.....-----C-T-----	TEEE-----T-----	288
HPV50	.....TTHTT..H-H-----TH-----	-----HHHHH	289
HPV60	.....TTTT-CHH-----HT-----CC-----	TTC-----HE---H	294
BPV1	.....-T-TTCCEE..E-----TT-E-----	EEEE-----H-----	297
BPV2	.....-TTTTCEE..ET-----T-C-----	EEEE-----	296
EEPV	.....-----E-----	EEEE-----	304
DPV	.....-----CCCC-----	CECC-----	304
BPV4	.....-----EE---TT-----	T-----T-----	289
HPV41	.....-----E-----TT-----	-----E---T-----	270
COPV	.....-EEET..T-----C-----	EE-----TTT-----	276
CRPV	.....--HHHHH---HHHHHHHHHH-----	E-----HH--	282
ROPV	.....-----C---CCC-----	-----EE-----	151
HPV1a	.....-----T-----	TT--T-----	292
HPV63	.....HHHHH..HHH...HHHHH-TT-E-----	EEEEETT-----	289
MnPV	CCCCCCCCCCC-E-----EE-----	EE-----	416

## E2 Appendix B

hpv_E2.allseqs.SOPM	ccccccc.CCCcC.....	Cc.hccccccCe	312
Gibrat_ALL_E2	EEEEEE-E---EE.....	EEEEEEHHH----H	308
Levin_ALL_E2	-----.	HHH-HHHHTT--C	308
DPM_ALL_E2	TTTTTT-T.T-T--..	-HC-HHHHH--C	308
SOPMA_ALL_E2	-----H..	HHH-HHHHH--C	308
Consensus_ALL_E2	-----.	HHH-HHHH--C	308
HPV54	-EEE.HTT.-TTT-..	---E--T-.--C	288
HPV32	.E-....	--T.T---	309
HPV42	EEEE-T-....	--H-....	313
HPV3	-----.	-----C	303
HPV28	--T-....-EEE..	-----E-	296
HPV10	EE-T-....	-----T.--	297
HPV29	--H.T-....	-----T.-E-	309
HPV61	E.ETT---.T---	-----	301
HPV2a	-----.	T-----C	311
HPV27	---.T-.T---T..	TT..T---.EE-	308
HPV57	-----.	-----T--.	303
HPV26	-----T..	-----EE-	298
HPV51	EEEE-T---.T..	-----EE-	283
HPV30	EE-----E..	ETCCCC--TT	299
HPV53	-.T--.T-....	CECCTCCCC--CCEEE---C	303
HPV56	-----T..	-----T	294
HPV66	--TTT--T---T..	-----T-E-	291
HPV18	T.---.T-....	EE..EE---.---	288
HPV45	HHE----.	EE..EE---.---	292
HPV39	HH--T----.---E..	EE..EETT--.E-	292
HPV70	.EE--.T---T-..	-E..EE---.E-	282
HPV59	-----TT..	-----	292
HPV7	-----HH.HHT-H..	T--.TTT---	296
HPV40	-----TT--..	E-----	291
HPV16	HEEE.HH-.T-TT-..	EE..EE-T.---	287
HPV35h	TEE--TT-....	-----T.	289
HPV31	EEE..HHH.----.	-----E..EE-T.---	294
HPV52	HEEE.ET---T-..	-----E..EEEEE--C	289
HPV33	HH-H.----T..	-----E..E---.---	275
HPV58	----T---.T-..	-----TT..	280
RhPV1	EEE--T-....	CC-----TT..	290
HPV6b	--EE---T-.TTT-..	-----T--.	286
HPV11	HHHEEH--T..TTT-..	-----TT--H..HE-	285
HPV44	-----E---TT-..	-----E---C	296
HPV55	-----TT-..	-----EE..E-	296
HPV13	-EEEEEE--TTTT..	-----EEEH..H-C	295
PCPV1	-----TT--..	-----C	295
HPV34	--T-....	-----E...E..E-	267
HPV19	T.-----CCCCCCC.CCEECECC..	CCCCEEECCCCCCC--CCHHHH--C	411
HPV25	-TEE.-----CCCCCCE.EEECCCCC..	CCCCEEECCCCCCC--CCHHHHHH--C	420
HPV20	--TT-C-----CTTCHEEE.ECTCCCC..	CTTEEEECCCCCTC-EC-HHHHHH--C	415
HPV21	---T-C-----CCCCCCEE.ECTCCCC..	CCCCCCCCCCCCCCCC--CEHHH--C	421
HPV14d	---T-C-----CCCCCTCC.CCCCCCCC..	CCCCCCCCCCCCCCCC--C-HHHH--C	401
HPV5	TT-TT---C-H--ECTTCHEEE.TECCCTC..	ETTEEEHHHHHHCTT--HH-HHHHTT--	432
HPV36	-----CCCCHEEC.TECCCTC..	CTTEEEHHCCCCCCC--C-HHHHH--C	427
HPV47	---EEE-C-----CECCCCCCC.CCCCCC..	CCTTEEECCCCCCCC--HH-HHHH--C	424
HPV12	--T-....CCCCCCC.CCCCCCCC..	CCCCCCCCCCCCCCCC--CCHHHHHH--C	412
HPV8	--TT-C-----CTCCCCHE.ECCCCCCC..	ECCEEHHCCCCCCC--C-HHHHH--C	416
HPV24	-----CCCCCTCC.CCCEECCTCCC..	TTCCCCCEECCCCCCC--CCHHHH--C	385
HPV15	-----TTCCCTC..TCCCCCCC..	HHHHHHHHCCCCHHHH--HHHHHT--C	374
HPV17	--TTT---TT-TTCTCCCTTCCCCHH..	HTHHHHHHCCCCCTT--H-HHHHH--C	370
HPV37	-TEEEE--EE---CCCCC..CT.TTECHHC..	TTCHHHHECCCCHHHH--HHH-T--C	372
HPV9	--T----.T---CTCCEEE.EECCTHC..	TTEEEEEECTTCCC--HH-HHHHTT--C	379
HPV22	TT--T-ET..TT-TT..	HHHHHHHHCTCCCTHHH--HHHHHTT--C	354
HPV23	EE--TT..EETT..	HHHHHHHHCTCCCTHHH--HHHHHT--C	349
HPV38	--E----.EECCCCCCC.CCCCCCCC..	CCEEEEEECCCCCCC--EEEEEE---C	359
HPV49	-----CCCCCCCC.CCCCCEEECCCCCCCCCCCCCCCC--TCC--	-----C	406
HPV4	----HH..H-HH..	HHHHHHHHHHHHHHHHHHHHHHHHHH--C	323
HPV65	-----E..EEE-H..	HHHTTHHHHHHH--H-T--C	323
HPV48	EET--EHE.E--T..	-----HHHHH--HHHTT--T	316
HPV50	H..HH..HHH..HHTH..	-----HHHHH--HHHTT--C	314
HPV60	HHHHHHHH--HCCHH..	-----HHH-HHTTT--C	324
BPV1	-----EE..	EECCCCCCC--EEEE..TT--	327
BPV2	-----E..	EECCCCCCC--H-EHHHTTT--	328
EEPV	-----EE..EE---CCCC..	-----EEE--T..C	332
DPV	-----E---CCCC..	-----E..EEEETTTTT--	333
BPV4	T.T----..	C.....CTTCCCCCTTCCHHHHHH--HHHHHH--T	327
HPV41	-----H..	CCCEEEHHHH--.TTE--T	300
COPV	--EE..EE-..	-----EEEEEHHHH--C	300
CRPV	EEEEEE--T----..	-----HEEEEEE---C	308
ROPV	-----.	CCCCCCCC--C-HH--	185
HPV1a	E.--E----..	-----CHHHHH--HHHHHT--T	319
HPV63	-----.	-----CCCHHH--HHHHHT--C	316
MnPV	---EEE---TT..	CCCCCCECCCCCCCCC-------T-E-	460



## E2 Appendix B

hpv_E2.allseqs.SOPM	ccccccceeeeccccccccceeccc...chc	393
Gibrat_ALL_E2	-HHHHHHHHHHHE---EE--CEE-...EC-	389
Levin_ALL_E2	-----HHHHH---S--EE- T--T...TTS	389
DPM_ALL_E2	H--HH---C--EE---EEE--C----C-	389
SOPMA_ALL_E2	--EE-HHHHHHE-----E---EE-...C-	389
Consensus_ALL_E2	--HH-HHHHHHE-----EE--C----C-	389
HPV54	--TT-E-----E--TEEEE---EE....E-	368
HPV32	T--TTE-----EEETCCT-T	395
HPV42	HHHHHHHHHHH---TT---CHHHHHHCCTT-	399
HPV3	-----TT-----E--C..-C-	384
HPV28	T--T-----E--T--EE---EE-E..EEE	377
HPV10	-----TE-----E--TE---EEE...EET	377
HPV29	T--T---HTC---THHHHHHTTEE...EEE	389
HPV61	---THHHHH---E-----EHHT--EE-H..H-T	383
HPV2a	--TT-----T---CCC---C..H-H	392
HPV27	TT--THH-----E---EEECCTE--H..EEE	389
HPV57	TT---H-----T---CTT-E-T..HCE	384
HPV26	--T-----E-----C-C----EEE	376
HPV51	HH-TTH-----T--HHCCCEH...H-E	359
HPV30	-----E-----E---E-TTC..TT	379
HPV53	-HTTH-C-H-----HE---HEEH..CE	385
HPV56	T--T---HHHE-----EEE-HTT-T	370
HPV66	T-TTT-----T---E-----EE....CH	370
HPV18	-----E-----EH-HHHH...-E-	366
HPV45	-H-TT-----E-----TE-CT-E....EET	369
HPV39	-TT-HH-----TTEEETTT--...-EE	371
HPV70	TT-TTHH-----TT--E-----...EEE	361
HPV59	--TTE-----TE-H-EE....EEE	371
HPV7	-----H-----E-----HHC---EEE...EET	376
HPV40	-TH-----E-----EEE...EET	371
HPV16	TT-TT---HHHH-----TEEEE---EE...H-T	366
HPV35h	-T-TT-----E---T-EEEC-----EE	368
HPV31	---T-HHC---E---T-ET--E...-EE	373
HPV52	---THE---E-----TEE---EE...EE-	369
HPV33	---HH-----C--EE....EE-	354
HPV58	--HHHE-----EECTT--...-	359
RhPV1	HH-HTHH-----EETTEEE-C--EE....EEE	367
HPV6b	TT--HH---T-----C--EE-EC.TTE	369
HPV11	---H---H-----T-C---EEEC.TTE	368
HPV44	T---HHH---T-----E--C---EE.EEE	378
HPV55	---HHH-----EEE-TT--EE.EEE	379
HPV13	---HHH-----HH---EEEEEE.EET	378
PCPV1	---TT---H---T--HH-HH-EETTT.TCT	378
HPV34	---HHC---TT-----ECCC---...-ET	346
HPV19	---CCC-----TT---C-H--H...TC-	494
HPV25	---CCCHH---TT-----H--H...TC-	503
HPV20	---HHH--E---TT---HHHHHH...T-T	498
HPV21	-----CC---TT---CCC--HH...H-T	504
HPV14d	-----HHHHC---TT---CHH--H...TCT	484
HPV5	-----HHHHHH---TT---HHHHHHH...H-H	515
HPV36	-----CC---E---T---CCC---...T--	510
HPV47	-----CCHHH---TT---HHHHH---...C-	507
HPV12	-----CCH---TT---HHHC--H...-H	495
HPV8	T-----CC---E---TT---CTC---...H-	499
HPV24	-----T-----E---TT---CHHHHH...HC-	468
HPV15	TTHHHH-----E---TT---HHHHHH...H-H	457
HPV17	--TTEE---TTT---T---CHC---...H--	453
HPV37	TT--HHE-----T---HHHHHH...H-H	455
HPV9	H-TT-E-----T---HHHH--HT....-E	462
HPV22	TT---HHHHHHHHHHE-EECTTT--...H-H	437
HPV23	TTTHHHHHHHHHH-E---E-TTTTT...-TH	432
HPV38	T-----E-----E-HC-H...TCE	442
HPV49	-----E---T---HHHH---...C-	489
HPV4	T---E---CCC---TTT-----E-T...CE	403
HPV65	TT---T-CCTE-T---EEE---C----C-	403
HPV48	T---TEE---E---T---E-C---T....-	397
HPV50	-T---HHHHHHH---T---E---E-H...H--	397
HPV60	---TT-----E---T---E-C-E...-C-	405
BPV1	-----TT---CC-EHHH...-H	411
BPV2	TT-----T---CC--EE-C...-E	412
EEPV	TT---E---C-----CHHEEEH..HC-	416
DPV	TT---EE-----CT---EEE..ECE	417
BPV4	TTTTT-----E---T---EHHTHHHH...H-H	409
HPV41	HHTHHHHHHHHH---TTHTHHCCHHH...HTT	388
COPV	-----E---H--EC-C---...CH	386
CRPV	--THH---C---T--HE--HEH...-C-	391
ROPV	-T-----CC-----E---T-T...-CE	268
HPV1a	H---T---HHH---E---TT-EE--C----TT-	402
HPV63	TT-T---HHH---E-----EE--TT-E...-CT	399
MnPV	-----C---C-E---T---E-C----C-	543

**Appendix C: Phenograms based on E2 Amino Acid Sequences**

Phenetic analysis is a form of cluster analysis that attempts to capture the relatedness of sequences irrespective of evolutionary pathways—that is to say the simple similarity of sequences. In the following two phenograms, E2 amino acid sequences over the NH<sub>2</sub>-terminal (about 200 aa) and COOH-terminal (about 90 aa) regions are compared using the PIMA program as described by Korber et al. (*J. Virol.* **68**:6730–674, 1994). The intervening hinge region, which is highly diverse, has been excluded from these analyses. The PIMA approach employs a hierarchical scoring scheme that allows for conserved substitutions in addition to identities. The abscissa records the similarity scores, whereas the ordinate merely records the number of sequences being compared. Sequences 44 and 55 are closely related (in both stretches of the E2 protein); in contrast, many of the sequences differ by as much as 70% using this scoring method—they are connected by nodes that are a small fraction of the score possessed by identical sequences, for example BPV-1 and BPV-2. The amino acid sequences cluster in these analyses as one would predict on the basis of phylogenetic classification—the groups of the A and B supergroups stay together in both fragments. One exception to this pattern is the RhPV1 sequence, which clusters with A9 PV sequences in general and in the C-terminal fragment but not in the N-terminal fragment. Otherwise, there appear to be no unexpected similarities nor unexpected dissimilarities as are seen with E4 sequences (Doorbar and Myers, Part III, appendix C).

